Using TraDIS to Probe the Model Organism *Escherichia coli*

by

Emily C. A. Goodall

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSPHY

Institute of Microbiology and Infection
School of Biosciences
College of Life and Environmental Sciences
University of Birmingham
September 2018

Abstract

Transposon-directed insertion-site sequencing (TraDIS) is a high-throughput method that couples transposon mutagenesis with short-fragment DNA sequencing. It is commonly used to identify the essential or conditionally essential genome of an organism, linking phenotype with genotype. A mini-Tn*5* transposon library was constructed in the model organism *Escherichia coli* K-12 strain BW25113. With a median distance between insertions of 3 bp, this is one of the most dense libraries published and provides a valuable tool for whole genome screening. Analysis of this library revealed subtle differences between the TraDIS method and the gold-standard gene-deletion method to determine the essential genome of an organism. This included, but was not limited to, the identification of transposon insertion bias, the boundaries of essential domains within a gene, and short essential domains within intragenic regions. Insertion bias was subsequently found to reveal the position of essential promoters and novel regulatory elements. This library was further used to probe the genomic requirement for survival in the presence of the clinically relevant antibiotic polymyxin B at sub-inhibitory concentrations. Among this data a gene of unknown function, *yhcB*, was identified and found to have a fundamental role in cell envelope biogenesis. The density of this library enabled identification of previously unreported genomic features. The results presented highlight the potential applications of TraDIS as a tool for a wide range of biological questions.

Tor and Fred, keep dreaming!

ACKNOWLEDGMENTS

Ian Henderson, I have had a fantastic time in your lab, thank you. Thanks for letting me have free reign in the lab, but for offering guidance when it is needed, even when it is not realised it is needed! Thanks for letting me work whatever hours I fancy, thanks for buying us burritos and coffee and booze, thank you for Ireland and America. Thank you for educating me on the ins and outs of a career in science and not just getting a Thesis done. Lastly, thank you for always having an open door and being understanding of those things outside of anyone's control.

To the PIs and professors I have been so lucky to have time and insight from, thank you Jeff Cole, Damon Huber and Pete Lund, it has been a real pleasure learning from you. I don't think I will ever forget my favourite question "but what does essential *mean*?". And Jeff, thank you for the considerable hours you have invested in trying to teach me to write, although I am still not sure if/whether I know the difference between 'if' and 'whether'...

Jack, as the original 'mini-boss' you get a special mention, and deserve an extra-special thank you. You have undoubtedly shaped my progress as a scientist and set the bar for what a scientist should be. (It is intimidatingly high). So many aspects of this project would not have come together without you, thank you. Most importantly, thank you for introducing me to bouldering!

To my 'Super Special Science Besties', Chris and Jess, I have had some fantastic times inside and outside the lab with you and my time at UoB would have been very different without you both! Thank you for all that you have taught me in the lab, but thank you also, together with Meek and Charles, for Tuesday come-dine, some horrific dancing over the years and always being the first to say yes to coffee/beer/Rymans. Sweden is one for the books.

To Georgia, 'Chief', and Nathaniel, thank you both for all the advice and help you continue to give me. You are both unfailing in finding the time to help me out no matter how busy you are. I value your expertise and more importantly your friendship.

Ash, Iain, Sara I would like to thank you also. I couldn't have asked for a nicer, more patient and understanding group of people to work with for my first experience of putting a publication together. To the rest of the Henderson lab and T101, for which there are so many of you to mention: Amanda, Emma, Sammi, Anna, JG, Kara, Dana, Karl, Rach, Tamar, Mo, Cesca, Fatima, Santosh and a guest appearance of Camilla, you all made the T101 lab a lively and entertaining place to work. Mat Milner, you are an absolute trooper and a real treasure to TraDIS with!

To my family, thank you Dad, for showing me how to put one foot in front of the other when the going gets tough. Thank you Tor, for showing me how to get the job done. Thank you Fred for showing me there is always more in the tank, no matter the time of day. And thank you Mum, for your never-ending provisions and perspective on what really matters. Thank you all for my commute chats over the years.

Saving the best 'til last, thank you Matt! Thank you for all the Saturdays you walked in to retrieve me from the lab, the Sundays you drove me in for overnights, the Mondays-Thursdays you listened to my failings and the Friday evenings when you came home and handed me Gin. Thank you for just being there throughout.

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Supplementary Tables

[a]These tables have been uploaded to the University of Birmingham eData repository (accession: DOI 10.25500/edata.bham.00000345) as they are too large to insert in the main body of the text.

Abbreviations

ACP	Acyl carrier protein
Amp^R	Ampicillin-resistance
ATP	Adenosine triphosphate
BACTH	Bacterial adenylate cyclase two-hybrid
BAM	β-barrel assembly machinery
bp	Base pair
CDP	Cytidine diphosphate
CDS	Coding sequence
CFU	Colony forming units
CL	Cardiolipin
COG	Cluster of orthologous groups
DNA	Deoxyribonucleic acid
ECA	Enterobacterial common antigen
EDTA	Ethylenediaminetetraacetic acid
FRT	Flp recombination target
GlcNAc	*N*-Acetylglucosamine
IFR	Insertion-free region
IIS	Insertion index score
IM	Inner membrane
Kan^R	Kanamycin-resistance
kb	Kilo base pair
LB	Luria Bertani
LPS	Lipopolysaccharide
Mb	Mega base pair
MCS	Multiple cloning site
MIC	Minimum inhibitory concentration
mRNA	Messenger RNA
MurNAc	*N*-Acetylmuramic acid
NGS	Next-generation sequencing
OD	Optical density

OM	Outer membrane
OMP	Outer membrane protein
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PGN	Peptidoglycan
PBS	Phosphate-buffered saline
PxB	Polymyxin B
qPCR	Quantitative PCR
RBS	Ribosome binding site
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RT	Room temperature
SDS	Sodium dodecyl sulfate
sRNA	Small ribonucleic acid
Tet^R	Tetracycline-resistance
TL	Transposon library
Tn	Transposon
tRNA	Transfer ribonucleic acid
TraDIS	Transposon-directed insertion-site sequencing
UIP	Unique insertion point
UDP	Uridine diphosphate
Und-P	Undecaprenyl-phosphate
UTR	Untranslated region
UV	Ultra violet

Chapter 1

Introduction

1.1 Introduction

Transposon-Directed Insertion-site Sequencing (TraDIS) is a high-throughput method that uses amplicon sequencing to identify the insertion site of a transposon within a library of transposon mutants. It is an excellent method for determining the essential genome of an organism, conditionally-essential genes, gene networks and novel genomic features overlooked or unidentifiable from other methods.

This thesis uses TraDIS to investigate the essential genes of the model organism *Escherichia coli* K-12 and the additional genomic features that can be revealed by such a screen. The technique was then applied to understand the role of conditionally-essential genes required for growth in the presence of polymyxin B, an antibiotic of significant clinical importance. The last chapter focuses on characterising a gene of unknown function identified in the polymyxin TraDIS screen that appears to have a fundamental role in correct cell growth and progression of the cell cycle, the characterisation of which may further our antimicrobial treatment capabilities.

The introduction chapter summarises the challenges posed by antibiotic resistance and describes some of the functions of the bacterial cell that are essential for viability, with a particular focus on the cell envelope as this forms the first line of defence to an antimicrobial compound. This chapter also reviews the advances in transposon mutagenesis and sequencing, and the applications of this powerful technology to answer a diverse range of biological questions.

1.2 Antibiotics and resistance

An antimicrobial drug is defined as a drug that kills or inhibits growth of a microorganism, termed bactericidal or bacteriostatic, respectively, when used to treat

bacteria. Antimicrobial resistance is when a microorganism is able to grow in the presence of a drug that would ordinarily inhibit growth of that species. Specifically, a bacterial strain is considered resistant to an antibiotic if the minimum inhibitory concentration of drug required to inhibit growth of the strain is higher than the designated clinical breakpoint for that species.

Bacterial resistance to antimicrobials is naturally occurring; resistance to assault enables microorganisms to live in a diverse range of harsh environments. Microorganisms competing for the same environmental niche have developed a multitude of antimicrobial strategies to gain a competitive advantage over neighbouring species. Similarly, bacteria must be able to withstand the antimicrobials they produce and have mechanisms to resist the antimicrobial substances produced by their competitors. Consequently, a large number of clinically available antibiotics have been derived from the natural products synthesised by microorganisms. This is especially true of the *Streptomyces* genus, which is said to be the source of around two thirds of all bacteria-derived antibiotics (Bérdy 2012).

While resistance to antibiotics is a natural phenomenon, the level of resistance an organism might have to a compound can vary. Under conditions of selective pressure the development of increased resistance is accelerated; something Alexander Fleming recognised in an interview in 1945, shortly after his Nobel acceptance speech, and warned against with the undisciplined use of antibiotic drugs (Anon 1945).

1.2.1 A history of antibiotics

People have been using organic remedies to treat infectious disease for millennia. There are reports of Ancient Egyptians treating skin infections with mouldy bread, a

1600 BC Greek king treating an infection with mouldy cheese and use of mouldy soya beans to treat infected wounds in China 3,000 years ago, despite lacking the knowledge of the causative infectious agent (Wainwright 1989). In a recent study, re-creation of an Anglo-Saxon recipe to treat eye infections, consisting of mulched onion, garlic and ox gall, among other things, was found to clear up to 90% of methicillin resistant *Staphylococcus aureus* in both a biofilm and an *in vivo* model (Harrison *et al.* 2015). The first use of a synthesised chemical compound to treat an infectious disease was the use of Arsphenamine (marketed as Salvarsan) to treat syphilis in 1909 (Williams 2009). The first mass-produced antibiotic was penicillin, a mould metabolite, discovered fortuitously by Alexander Fleming in 1928 (Fleming 1929).

Shortly afterwards began the 'Golden Age' of antibiotics where a worldwide hunt for new antibiotics began (Davies 2006). The search for antibiotics often centred on compounds synthesised by bacteria themselves, which drove an interest in acquiring and characterising new species of bacteria. For example, the company Eli Lily capitalised on the far-reaching movements of Christian missionaries asking them to send home soil samples from their travels (Gould 2016). One such trip to Borneo unearthed a strain of *Streptomyces orientalis,* which was found to synthesise vancomycin. Although initially dubbed "Mississippi mud" because of its pre-purification colour, the compound was given the name "vancomycin", derived from the word "vanquish" (Gould 2016; Levine 2006).

The period from the 1940's until the end of the 1970's resulted in the discovery of most classes of antibiotics in use today. A timeline of the discovery of some of the major classes of antibiotic is shown in figure 1.1. The introduction of each new antibiotic ultimately led to the concomitant emergence of resistance to that antibiotic

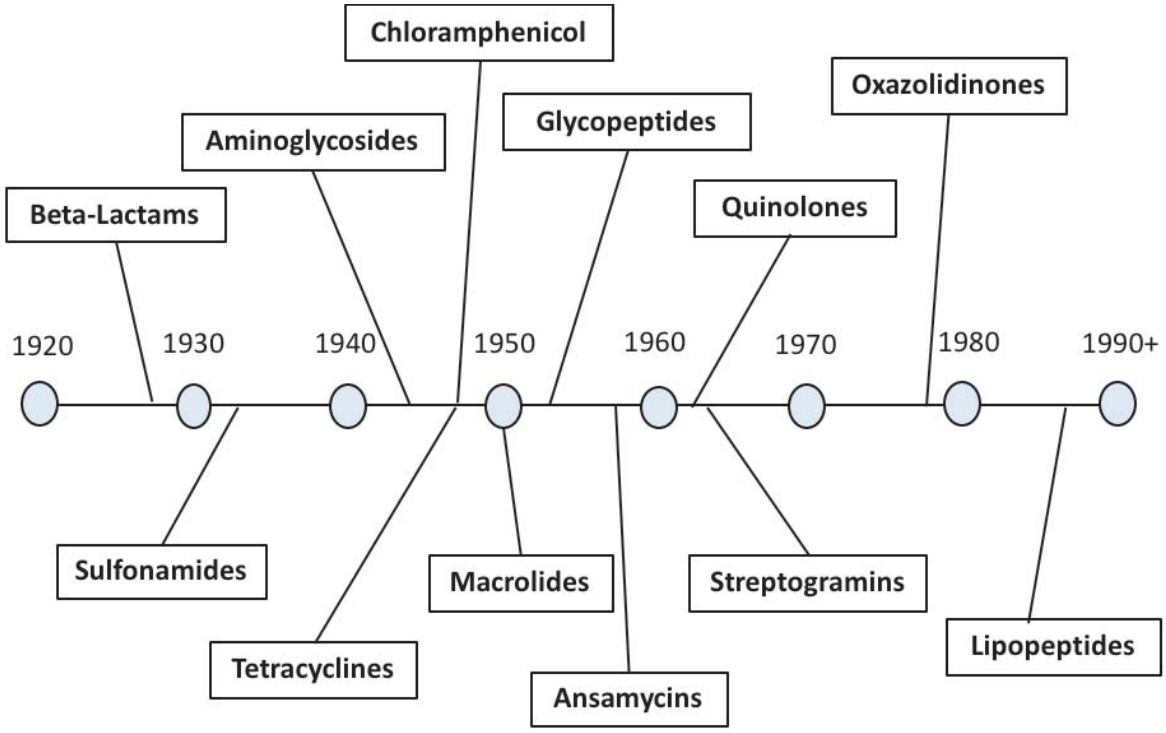

Figure 1.1 Timeline of the discovery of the major antibiotic classes

The discovery of the first drug of each major class of antibiotics is indicated on the timeline by the antibiotic class name, from the period spanning 1920-1990.

in a clinical setting (Lewis 2013). One means of combatting this was to develop existing drugs to improve the spectrum of activity and potency. An example of this is Cefepime, a 4th generation cephalosporin, belonging to the β-lactam class of antibiotics. Unlike earlier generations, the 4th generation of cephalosporins are zwitterionic, which improves their ability to cross the OM in Gram-negative bacteria. Specific side chains of Cefepime also confer additional protection from β-lactamases. However, with a lack of any new drugs (referred to as the "discovery void"), the increased use of antibiotics and the selective pressure this has imposed on bacteria, the number of antibiotic-resistant infections has steadily increased as a result (Silver 2011).

1.2.2 The clinical significance of antibiotic resistance

Antibiotic resistance is described by the World Health Organisation (WHO) as "one of the biggest threats to global health, food security, and development today" (WHO 2018). Bacterial infections that cannot be cleared by currently available drugs are often associated with worse patient outcome and are increasing globally (Blair *et al.* 2015). Antimicrobial resistance reportedly accounts for around 23,000 and 25,000 deaths annually in America and Europe, respectively, and has been calculated to cost the US between $21 and $34 billion annually (Davies 2011; WHO 2014).

The most prevalent cases of resistant hospital-acquired (nosocomial) infections are caused by the ESKAPE pathogens, an acronym derived from the species within this group: *Enterococcus faecium*, *S. aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumanii*, *Pseudomonas aeruginosa* and *Enterobacter* species (Rice 2010). Collectively, these species are considered to account for the majority of resistant

nosocomial infections. As many as 60% of *E. faecium* isolates are estimated to be resistant to vancomycin, between 65 and 85% of *S. aureus* isolates have been identified as resistant to penicillin G, and the prevalence of methicillin resistant strains of *S. aureus* is reportedly around 25%, increasing to 50% in some areas (Santajit & Indrawattana 2016). β-lactamases, which are especially common among the Gram-negative *Enterobacter* species, have been identified in resistant isolates globally (Shaikh *et al.* 2015). The New Delhi metallo-β-lactamase, a broad spectrum β-lactamase, has been reported in over 20 countries (Velkov *et al.* 2013). Without the development of new antibiotics, or treatment strategies, there is a risk of entering a post-antibiotic era where routine operations such as a hip replacement will carry a much higher risk of complication.

1.2.3 Antibiotic resistance mechanisms

There are many ways in which an organism can become resistant to an antibiotic, but these can generally be grouped into two categories: intrinsic resistance and acquired resistance (Blair *et al.* 2015). Intrinsic resistance is the innate ability of an organism to withstand assault, for example via drug efflux, or an impervious membrane providing barrier protection. These properties are inherently encoded on the genome. In contrast, acquired resistance involves genetic modification to increase resistance either by mutation or horizontal gene transfer.

There are four general mechanisms that can confer resistance to an antibiotic. 1) The inability of a compound to access its target, the presence of target homologues, or the absence of a target altogether. 2) Reduced concentration of the compound within in the cell, either via decreased barrier permeability so less compound can

enter the cell, or via increased efflux, so more compound is ejected from the cell. 3) Modification of the compound target, to prevent drug activity, and 4) Modification of the compound. Examples of each of these are presented below.

(1) The antibiotic triclosan targets the enoyl-acyl carrier protein reductase (FabI in *E. coli*), however, *P. aeruginosa* is less susceptible to triclosan because it has two enoyl-reductase isozymes, FabI and FabV, the latter of which confers resistance to triclosan (Zhu *et al*. 2010).

(2) A common resistance mechanism that lowers the internal concentration of an antibiotic is the increase in expression of the AcrAB-TolC efflux pump (Okusu *et al*. 1996). The AcrB pump spans the inner membrane and crosses the periplasm, pumping molecules from the inner membrane or periplasm out through the TolC channel, powered by the proton motive force. The AcrAB efflux pump is effective because it can efflux a wide range of molecules (Blair & Piddock 2009), although the activity and compound specificity can be modified by the small inner membrane protein AcrZ (Hobbs *et al*. 2012). Upregulation of AcrAB expression results in increased efflux activity, mitigating the damaging effects of toxic compounds (Keeney *et al*. 2007; Keeney *et al*. 2008).

(3) An example of target modification to confer resistance is the adaptation of the surface exposed lipopolysaccharide (LPS) in Gram-negative bacteria in response to exposure to polymyxins (Olaitan *et al*. 2014). For example, substitution of the side-branch phosphate groups for phosphoethanolamine, or 4-amino-4-deoxy-L-arabinose, decreases the overall negative charge of the lipid A moiety (Gunn *et al*. 1998). This reduces the binding efficiency of polymyxins with lipid A, which in turn reduces the efficacy of polymyxins.

(4) Lastly, arguably the most well-known example of antibiotic modification that confers resistance is the hydrolysis of the β-lactam ring within β-lactam antibiotics (e.g. penicillin) by β-lactamases, rendering the antibiotic ineffective. The first β-lactamase was isolated in 1940 and found to degrade penicillin (Abraham & Chain 1940). The identification of β-lactamases within both clinical and environmental isolates is widespread. The activity of β-lactamases for specific β-lactams varies; some β-lactamases are able to hydrolyse a wide range of antibiotics, such as the extended-spectrum β-lactamases (Shaikh *et al.* 2015). These are especially problematic when encoded on plasmids or in proximity to mobile genetic elements as they are easily disseminated via horizontal gene transfer to other strains and species of bacteria. The widespread dissemination of β-lactamases among clinical isolates has serious implications for the effective treatment of complex infections (Blair *et al.* 2015).

Understanding the cell response to an antibiotic is important because it can reveal pathways that have implications for the effectiveness of clinical treatment. For example, a recent high-throughput screen that reported the synergistic effects of colistin and clarithromycin, also identified that a food additive, vanillin, increased the expression levels of the transporter MdfA, resulting in increased sensitivity of the cell to spectinomycin (Brochado *et al.* 2018). However, compounds can also have antagonistic effects, for example vanillin also induced increased expression of the efflux pump component AcrA, resulting in an increase in the concentration of ciprofloxacin or chloramphenicol required to inhibit cell growth (Brochado *et al.* 2018). Understanding synergistic relationships between antimicrobial compounds is important because it has implications for the use of antibiotics at lower doses, which

might serve to improve host-toxicity, and can further the use of otherwise ineffective antibiotics.

The search for new antibiotics, antibiotic targets or methods of treatment is of fundamental importance given the increasing levels of antimicrobial resistance identified in clinical isolates and the associated increase in mortality. A detailed understanding of bacterial cell physiology can result in the identification of biochemical pathways or proteins that can be targeted by antibiotics. However, unless an antibiotic specifically targets surface exposed molecules, often, the first barrier to an antimicrobial compound is the cell envelope.

1.3 The Gram-negative bacterial cell envelope

Most bacteria are fundamentally classed into one of two distinctive sub-groups: Gram-positive and Gram-negative bacteria (Silhavy *et al.* 2010). The key difference between the two is the structure of the cell envelope. Gram-positive bacteria have an inner membrane lipid bilayer (IM) and a complex outer cell wall, while Gram-negative bacteria are encapsulated by three successive layers: an IM, a thin peptidoglycan cell wall and an asymmetric outer membrane (OM) lipid bilayer (Fig. 1.2). The presence of three layers in Gram negative bacteria was first observed by Kellenberger and Ryter by electron microscopy (1958). The three layers were later defined as the IM, cell wall and OM from thin-layer electron micrographs and subsequent analyses (Glauert & Thornley 1969; Bladen & Mergenhagen 1964).

The cell envelope serves both to keep cellular components in, and toxic compounds out. The two-membranes provide an impressive barrier system that conveys selective permeability as the OM and IM have opposite permeability

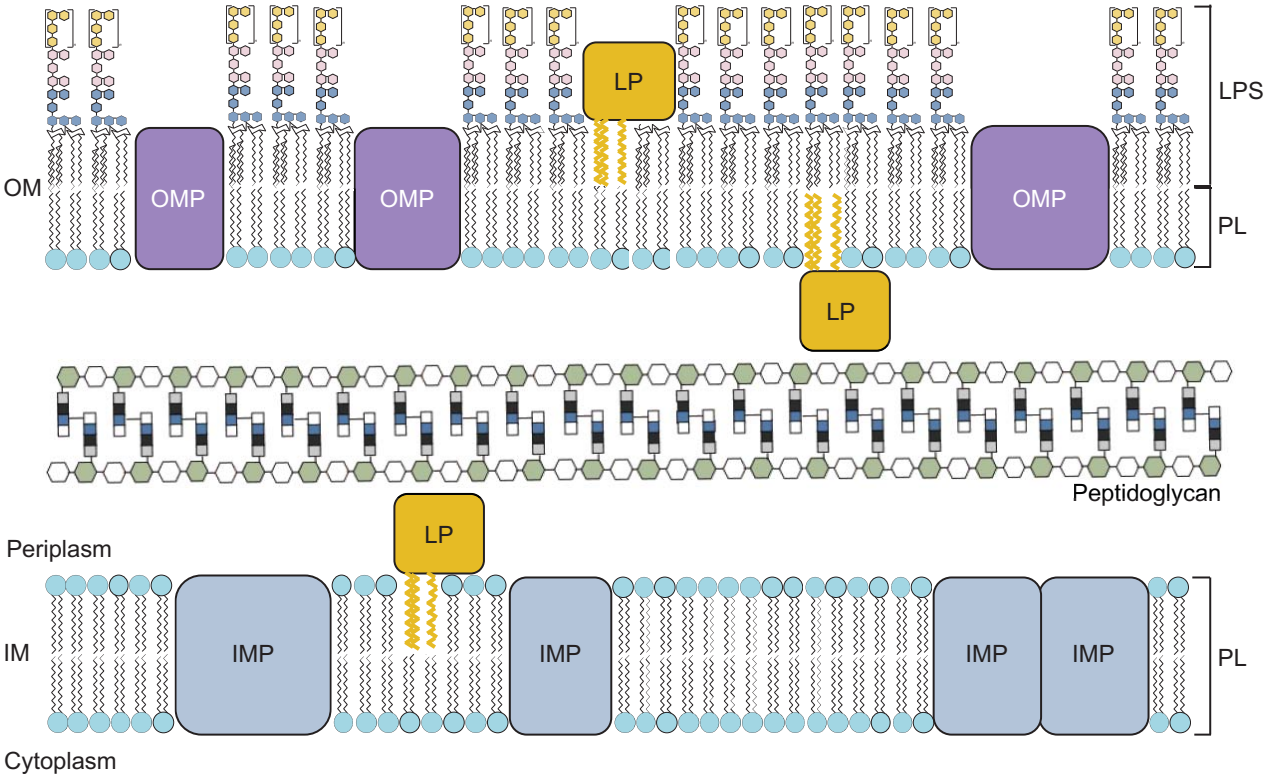

Figure 1.2 The Gram-negative bacterial cell envelope

The cell envelope of Gram-negative bacteria is comprised of three layered structures, the inner membrane (IM), peptidoglycan cell wall and the outer membrane (OM). The asymmetric OM is made up of phospholipids (PL; blue) on the inner leaflet, and lipopolysaccharide (LPS) on the outer leaflet. Outer membrane proteins (OMPs; purple) are β-barrel proteins embedded within the membrane, and lipoproteins (LP; yellow) are anchored by tri-acyl chains inserted into the membrane. The IM is a phospholipid bilayer with α-helical integral membrane proteins (IMP; light blue). Sandwiched between the two layers is the peptidoglycan cell wall, made up of disaccharide repeating units crosslinked by a pentapeptide, the terminal D-alanine of which is cleaved in mature peptidoglycan.

properties. The IM allows diffusion of small hydrophobic molecules, which are excluded from crossing the OM by the charge presented by the surface exposed LPS. While the OM allows diffusion of small hydrophilic molecules through integral porins. The additional OM layer in Gram-negative bacteria forms a strong permeability barrier conferring further resistance to toxic antimicrobials in comparison to most Gram-positive bacteria. Therefore, this layer can enable survival in a range of hostile environments such as those encountered during infection and host colonisation. Understanding how this complex envelope is synthesised and maintained is instrumental for understanding how to disrupt its function and render the bacterium susceptible to otherwise ineffective treatment.

1.3.1 The inner membrane and integral proteins

The IM is a symmetrical bilayer of phospholipids with imbedded and associated proteins. It provides a hydrophobic barrier that separates the cytoplasm from the periplasm, mediating selective permeability between the two compartments. Small, hydrophobic molecules can freely diffuse across the membrane, but larger or hydrophilic molecules require active transport. The compartmentalisation of the cell and ability to control directional movement of molecules across the membrane is an invaluable resource that can be harnessed for generating energy (Harold 1972).

The proteins of the IM can be divided into three groups: Integral proteins, those with at least one hydrophobic α-helical transmembrane domain imbedded within the membrane; lipoproteins, imbedded into the IM via three acyl chains, and peripherally associated proteins, which are typically hydrophilic and interact directly with the membrane surface or other IM proteins (Papanastasiou *et al.* 2013). Integral

IM proteins have a range of important biological functions including: metabolism and energy generation; protein secretion; biogenesis of components of the cell envelope; transport of small molecules; cell signalling; maintenance of the cell structure; cell division; and drug efflux, among others (Fig. 1.3; Bendezú *et al*. 2009; Morgenstein *et al*. 2015; Dalbey & Chen 2004; Hobbs *et al*. 2012; Kundig *et al*. 1964). As the site of ATP synthesis and synthesis of most structural components of the cell envelope, such as phospholipids, LPS, and peptidoglycan, the IM is arguably the powerhouse of the cell. In *E. coli*, around a quarter of the total genes are reported to encode IM proteins (Luirink *et al*. 2012).

Proteins of the IM are supported by a fluid bilayer of phospholipids. In *E. coli*, the three most common phospholipids of the IM are the zwitterionic phospholipid phosphatidylethanolamine (PE), and the anionic phospholipids phosphatidylglycerol (PG) and cardiolipin (CL). Under normal growth conditions PE makes up around 75% of the membrane, while PG and CL make up approximately 20% and 5% of the membrane, respectively (Rowlett *et al*. 2017). The anionic phospholipids have both been associated with negative curvature of cell membrane (Renner & Weibel 2011). Cardiolipin localises to the poles and sites of septum formation and is thought to have a role in maintaining membrane integrity under conditions of heightened physiological stress (Mileykovskaya & Dowhan 2000). Furthermore, upon entry into stationary phase the cell undergoes membrane remodelling and increases the overall proportion of cardiolipin (Hiraoka *et al*. 1993).

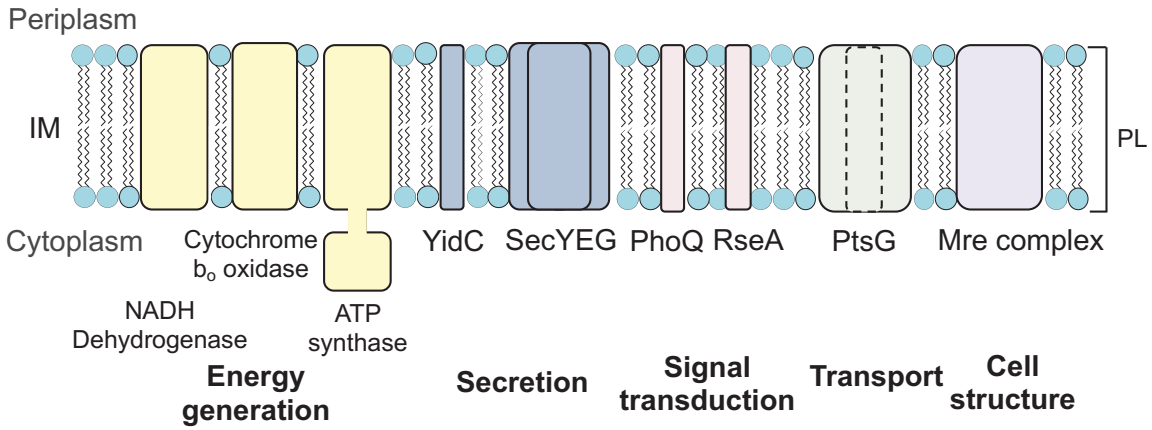

Figure 1.3 Representative integral inner membrane proteins

Examples of some of the many integral IM proteins are presented above, coloured according to function (Energy generation, yellow; Secretion, blue; Signal transduction, red; Transport, green; Cell structure, purple). Abbreviations: IM, inner membrane; PL, phospholipids.

1.3.2 The outer membrane

In *E. coli*, the OM is essential for survival. It is an asymmetrical bilayer with a phospholipid inner leaflet and LPS outer leaflet forming a selectively permeable barrier. The asymmetric nature of the membrane was determined by two simultaneous studies, the first showing that phosphatidyl-ethanolamine was not modified by externally provided phospholipases and is therefore not surface exposed (Kamio & Nikaido 1976). The second showed, through immunoelectron microscopy, that ferritin-conjugated antibodies directed against LPS label the outer leaflet of the membrane (Mühlradt & Golecki 1975).

Interspersed throughout the OM are a range of embedded β-barrel proteins, called outer membrane proteins (OMPs), and membrane-anchored lipoproteins. β-barrel proteins are trafficked to the OM by chaperone proteins SurA, Skp and DegP (discussed in section 1.3.6) and are inserted into the OM by the β-barrel assembly machinery (BAM) complex (Fig. 1.4). Lipoproteins are trafficked to the OM via the localization of lipoproteins (Lol) pathway (Matsuyama *et al.* 1995).

The structure of the BAM complex consists of the β-barrel protein BamA and four lipoproteins BamB-E (Iadanza *et al.* 2016; Bakelar *et al.* 2016; Han *et al.* 2016). Both BamA and BamD are essential. The lipoproteins BamB-E are thought to mediate β-barrel protein substrate specificity (Mahoney *et al.* 2016). The BAM complex is thought to insert proteins into the OM via lateral gating and localised membrane thinning, whereby the β-barrel of BamA opens in a clasp-like fashion allowing newly folded OMPs access into the OM (Noinaj *et al.* 2014). Cysteine-crosslinking to lock the BamA gate shut resulted in a decrease in BAM activity *in vitro* and was lethal *in vivo* under oxidising conditions (Iadanza *et al.* 2016; Noinaj *et al.* 2014).

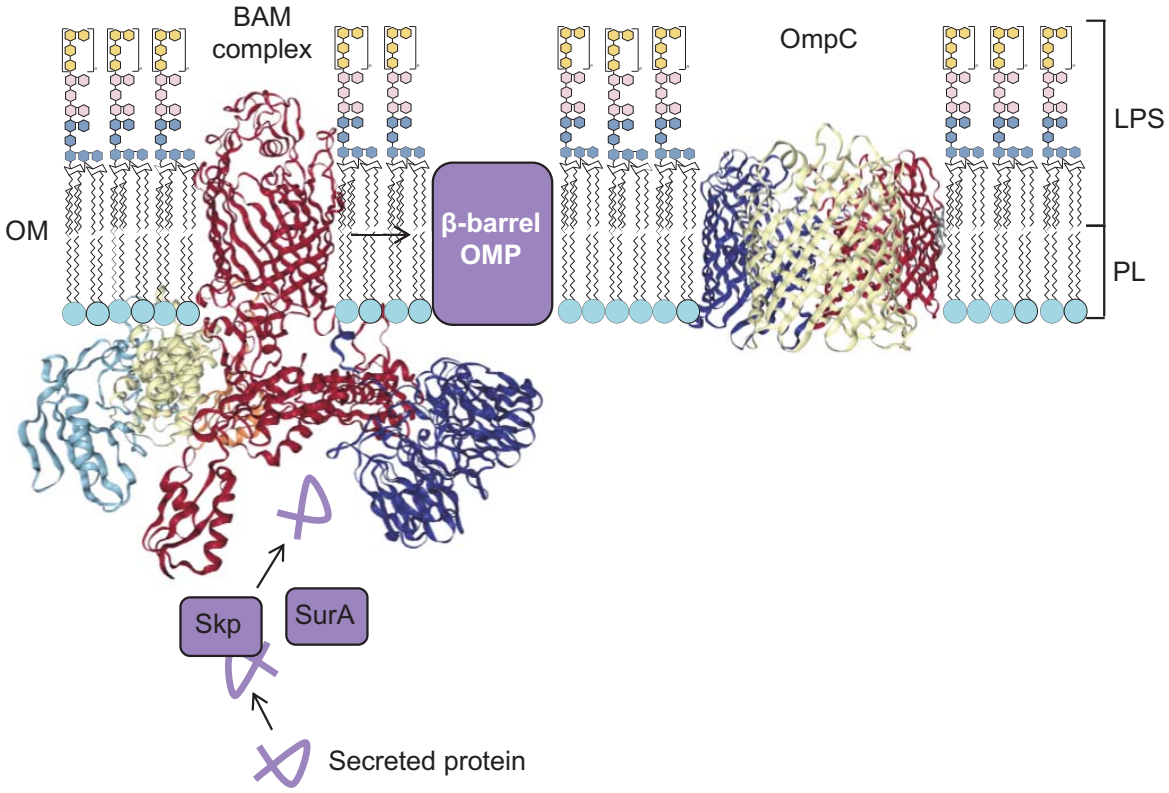

Figure 1.4 The outer membrane and β-barrel assembly machinery

The outer membrane (OM) is an asymmetrical bilayer made up of lipopolysaccharide (LPS) and phospholipids (PL) on the outer and inner leaflet respectively. Imbedded within the membrane are β-barrel outer membrane proteins (OMPs), which are assembled and inserted by the β-barrel assembly machinery (BAM) complex. β-barrel proteins destined for the OM are trafficked to the BAM complex by chaperone proteins Skp and SurA (purple). The structure of the BAM complex is taken from Iadanza *et al.* (2016). Components are coloured according to protein (BamA, red; BamB, dark blue; BamC, pale blue; BamD, cream; BamE, orange). The trimeric β-barrel structure of the outer membrane protein OmpC is adapted from Baslé *et al.* (2006). The structure is coloured according to individual OmpC monomers.

Once inserted into the OM, β-barrel OMPs have a barrel-like structure of usually 8-22 β-strands (Fig. 1.4; Schulz 2002). In *E. coli* the most abundant OMPs are the trimeric porins OmpC and OmpF (Masi & Pagès 2013). Porins form channels in the OM allowing the passive diffusion of small hydrophilic molecules <600 Da into the cell (Schulz 1993; Cowan *et al.* 1992). However, this also allows access of small, hydrophilic antibiotics such as the β-lactams and fluoroquinolones (Ceccarelli & Ruggerone 2008). Other OMPs have a role in the efflux of compounds, such as TolC, or substrate secretion, such as the β-barrel proteins of the Type I-VI secretion systems (Green & Mecsas 2016). OMPs can also have roles in degradation, such as PagP, which transfers an acyl chain from phospholipids to lipid A in the OM resulting in hepta-acylated LPS (Bishop *et al.* 2000).

Lipoproteins also have a diverse range of functions. The most abundant lipoprotein, with an estimated 750,000 copies, is Lpp. Lpp covalently links the OM to the cell wall (Braun & Rehn 1969). Amazingly, an *lpp* deletion strain is viable, but the weakened envelope results in increased blebbing, leakage of periplasmic proteins and increased sensitivity to envelope stress (Hirota *et al.* 1977; Suzuki *et al.* 1978). A recent study showed that the length of Lpp is the main determinant of the periplasmic width (Asmar *et al.* 2017). Lipoproteins are also found in every major cell envelope biosynthesis complex, highlighting the important functional role of OMPs.

The proteins of the OM are surrounded by a phospholipid-LPS bilayer. The overall charge of LPS is negative. In the presence of divalent cations, the LPS molecules pack tightly together further reducing the permeability of the membrane (Schneck *et al.* 2010). The lateral movement of LPS molecules, coupled via Ca^{2+} bridges, is highly ordered and slower than the random lateral movement of

phospholipids in the inner leaflet of the OM (Piggot *et al*. 2011). The lateral motion of LPS mirrors that of integral OMPs and is therefore thought to have a role in the movement of proteins within the membrane (Hsu *et al*. 2017).

Loss of LPS is lethal among most Gram-negative bacteria such as *Salmonella enterica*, *E. coli* and *P. aeruginosa*, however this is not true of all Gram-negative bacteria. Rare mutants of *Acinetobacter baumannii*, *Moraxella catarrhalis* and *Neisseria meningitidis* without LPS have been isolated (Simpson *et al*. 2015; Peng *et al*. 2005; Steeghs *et al*. 1998). In *A. baumannii,* loss of LPS can even confer a selective advantage when mutants are exposed to cationic peptide antibiotics such as colistin or polymyxin B, as the loss of LPS removes the outermost antibiotic target (Moffatt *et al*. 2010). Conversely, when the quantity or stability of OM LPS is decreased in *E. coli*, this results in an increase in surface exposed phospholipids, thought to form rafts or patches, and causes an increase in sensitivity to hydrophobic molecules (Simpson *et al*. 2015).

The surface exposed LPS is also a potent toxin that triggers a massive immune response in humans (Raetz & Whitfield 2002). This response is mediated via interaction with the TLR4 (toll-like receptor 4) receptor, which is surface exposed on monocytes and macrophages (Poltorak *et al*. 1998). Activation of TLR4 results in downstream activation of the NF-κB transcription factor and ultimately results in the release of pro-inflammatory cytokines. A response that is beneficial following a localised exposure, but detrimental in a systemic infection (Raetz & Whitfield 2002).

1.3.3 Lipopolysaccharide biosynthesis

LPS is an amphipathic molecule made up of three distinct structural domains: lipid A, core oligosaccharide and O-antigen (Fig. 1.5). The lipid A moiety anchors the structure in to the membrane. Attached to the lipid A moiety is the core oligosaccharide region, which can be further subdivided into an inner- and outer-core. The outer oligosaccharide, or O-antigen is the outermost surface-exposed component. The O-antigen chain length in *E. coli* can vary from 5 to >100 units but is often lacking altogether from laboratory strains, as is the case with *E. coli* K-12 strains (Osawa *et al*. 2013). The lipid A domain and Kdo (keto-deoxyoctulosonate) moieties of the inner core are essential in *E. coli* and the bare minimum components required for viability (Brabetz *et al*. 1997). There are approximately 10^6 lipid A molecules in a single *E. coli* cell (Raetz & Whitfield 2002).

Although exposed at the outer leaflet of the OM, LPS is synthesised at the inner leaflet of the IM. LPS molecules are then transported to the OM by the Lpt machinery (discussed in section 1.3.6). The first step of LPS biosynthesis is catalysed by the acyl transferase LpxA (Fig. 1.6). LpxA acylates UDP-*N*-acetylglucosamine with an acyl chain extracted from 3R-3-hydroxytetradecanoyl-ACP (Anderson & Raetz 1987).

The two input substrates of the reaction are both key compounds that are used in the synthesis of multiple components of the cell envelope. UDP-*N*-acetylglucosamine is used as the starting point for peptidoglycan (PGN) biosynthesis, while (R)-3-hydroxytetradecanoyl-ACP is used in palmitate and diacylglycerol biosynthesis. Further acylation steps by LpxC and LpxD form UDP-diacylglucosamine, which is dephosphorylated by LpxH to form lipid X (Fig. 1.6). UDP-diacylglucosamine and lipid X are combined by LpxB to form disaccharide-1-phosphate which is then

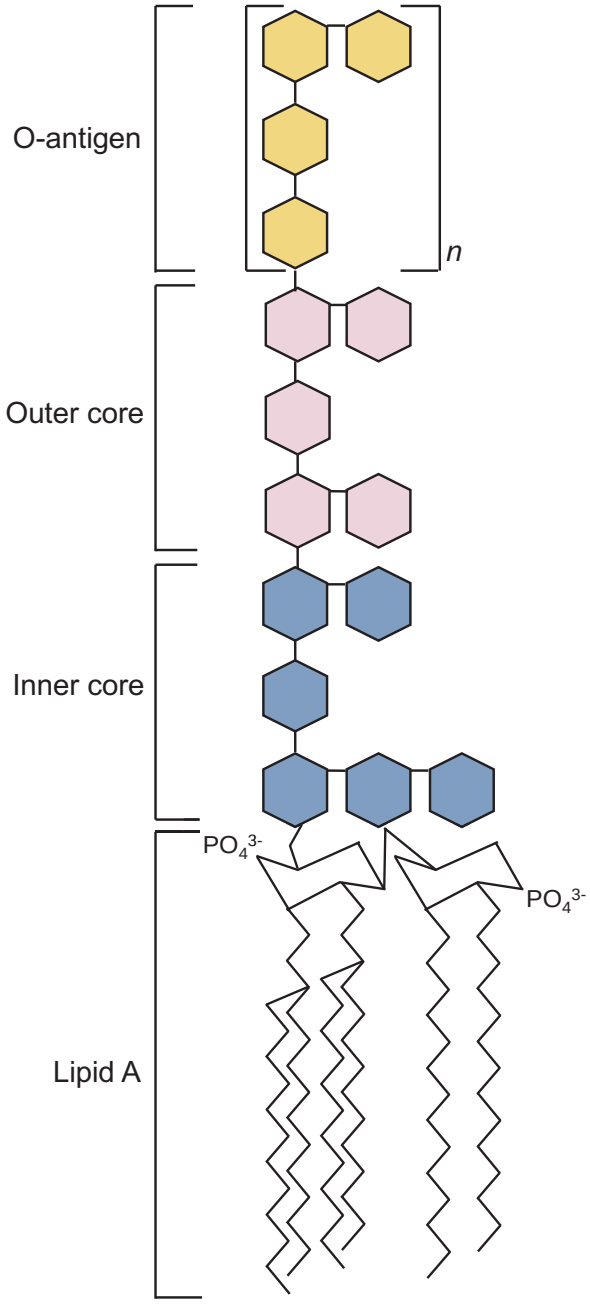

Figure 1.5 The structure of lipopolysaccharide

A schematic of the structure of LPS coloured by structural domains: Inner core (blue), outer core (pink) and O-antigen (yellow). The repeating units of O-antigen are indicated by a square bracket. However, the O-antigen portion is absent altogether in K-12 strains. Inner core side chains such as phosphate and pyrophosphorylethanolamine are not included for simplicity.

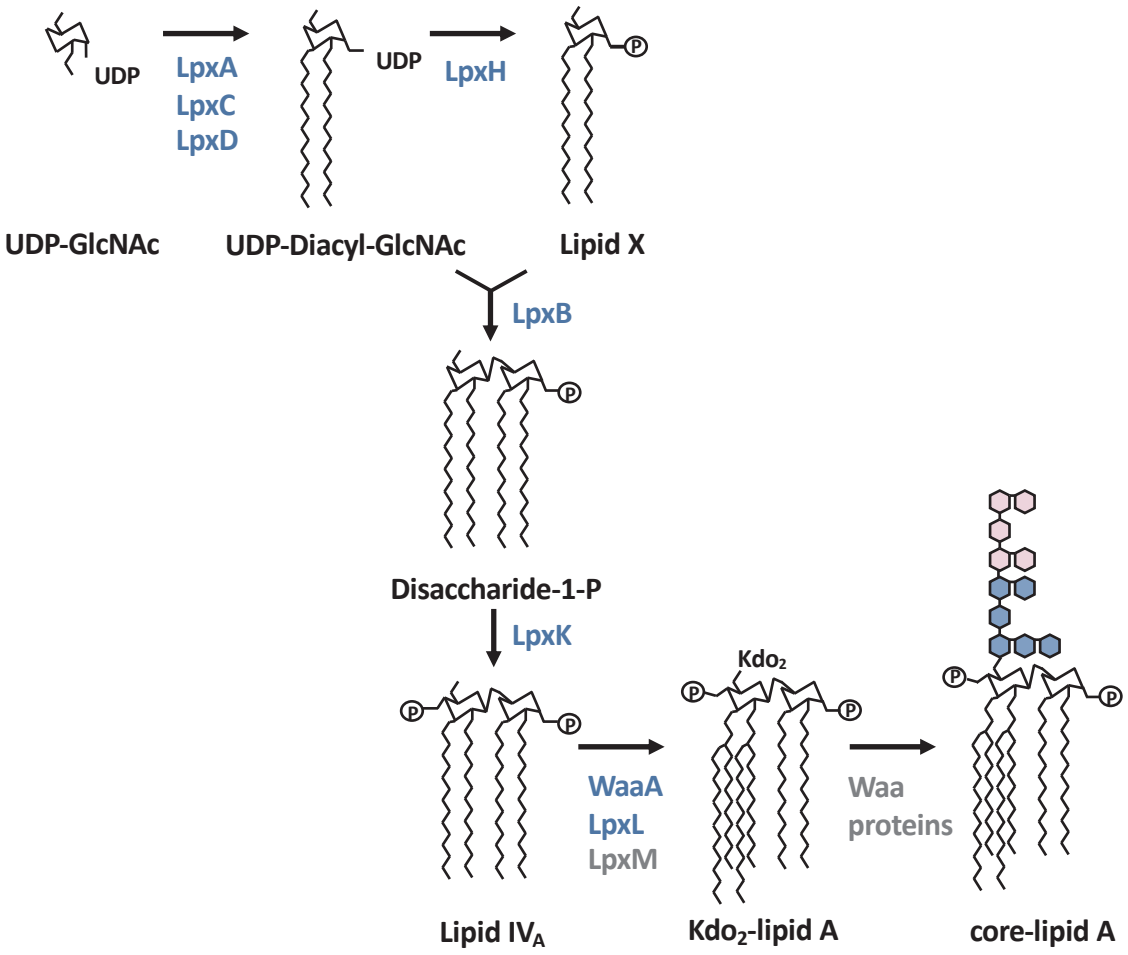

Figure 1.6 Lipopolysaccharide biosynthesis

Intermediates of the core-lipid A biosynthesis pathway. Essential enzymes of the LPS biosynthetic pathway are labelled in blue, non-essential enzymes are in grey. The inner core is coloured in blue, and the outer core in pink. (Abbreviations: GlcNac, *N*-Acetylglucosamine; Kdo, 3-deoxy-D-manno-octulosoni acid; P, phosphate; UDP, uridine diphosphate)

phosphorylated by LpxK to form lipid IV_A. Two Kdo domains are sequentially transferred onto lipid IV_A by the Kdo transferase WaaA. The Kdo_2-lipid $1V_A$ complex is then sequentially acylated by LpxL followed by LpxM to form Kdo_2-lipid A (hereafter lipid A; Raetz & Whitfield 2002). All enzymes except LpxM are essential.

Enzymes encoded by the *waa* operon sequentially add the core sugars onto lipid A at the cytoplasmic face of the IM. The inner core is highly conserved within species but the outer core exhibits minor variations. There are at least 5 outer core types of *E. coli*, R1, R2, R3, R4 and K-12 (Raetz & Whitfield 2002). Core-lipid A is then transferred across the IM by the flippase MsbA (Zhou *et al.* 1998). In *E. coli* strains with O-antigen, the O-antigen is independently synthesised and ligated onto core-lipid A by the ligase WaaL at the periplasmic face of the IM (McGrath & Osborn 1991). However, in *E. coli* K-12 strains, O-antigen is not synthesised because of one of 2 mutations in the *rfb* operon (Liu & Reeves 1994; Stevenson *et al.* 1994). The final LPS structure is then transported across the periplasm by the Lpt system (Polissi & Sperandeo 2014; Simpson *et al.* 2015).

1.3.4 Phospholipid biosynthesis

Phospholipids, generally, are composed of a glycerol backbone with two fatty acid chains and a phosphate group, with a variable head group. As such, the starting point of phospholipid biosynthesis is glycerol-3-phosphate (Fig. 1.7). The first committed step in phospholipid biosynthesis is catalysed by PlsB, an acyltransferase that adds the first acyl chain to glycerol-3-phosphate to form 1-acyl-*sn*-glycerol-3-phosphate (Larson *et al.* 1980). The 1-acyl-*sn*-glycerl-3-phosphate molecule is further acylated by an additional acyltransferase, PlsC, which catalyses the addition of another acyl chain to

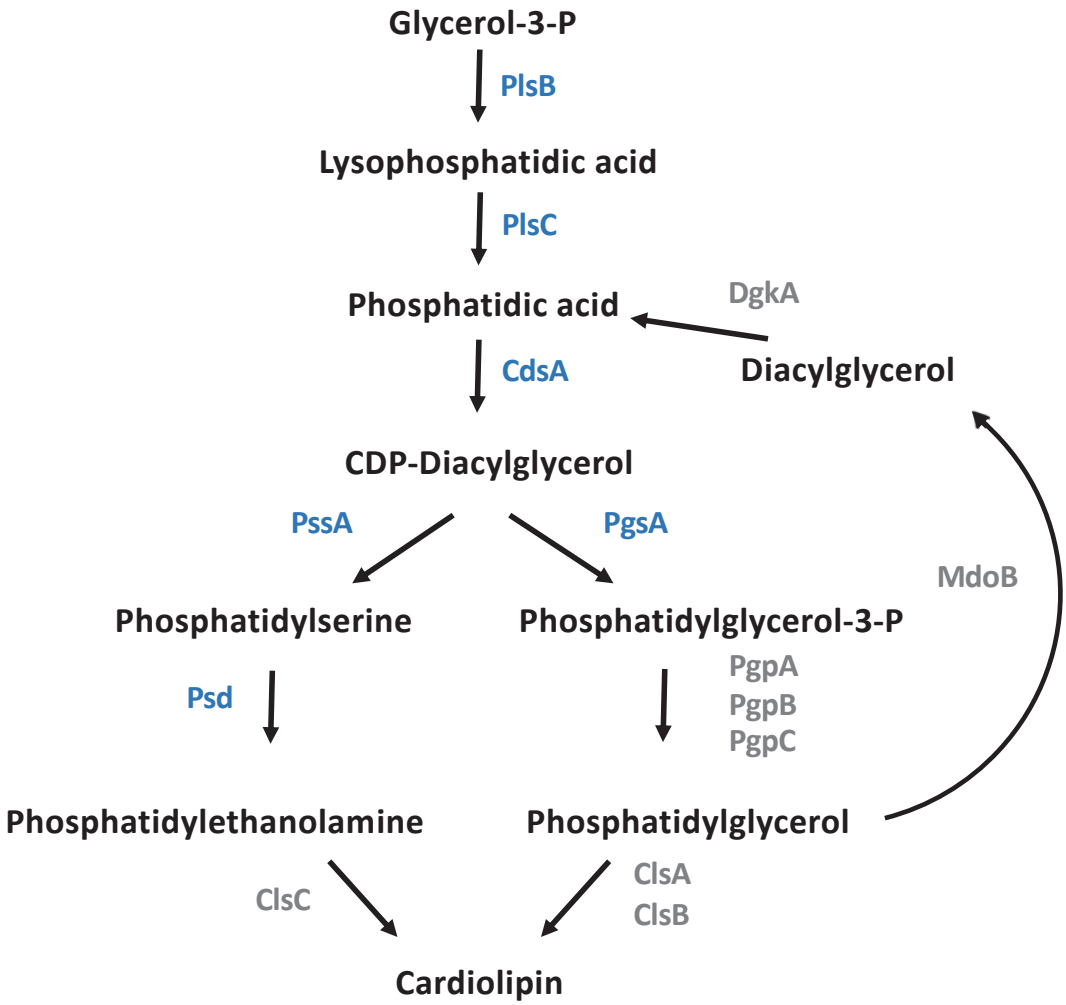

Figure 1.7 Phospholipid biosynthetic pathways

Enzymes are shown in next to the process they mediate. Enzymes that are essential in *E. coli* K-12 are shown in blue, non-essential enzymes are shown in grey. (Abbreviations: Glycerol-P: glycerol-3-phosphate; CDP-Diacylglycerol: cytidine diphosphate-diacylglycerol; phosphatidylglycerol-3-P: phosphatidylglycerol-3-phosphate)

form 1,2-diacyl-*sn*-glycerol-3-phosphate. PlsB and PlsC are both IM proteins and thought to be in close proximity to each other (Kessels *et al*. 1983). Acyl chains in both reaction steps can be taken from either acyl-ACP (acyl carrier protein) or acyl-CoA (coenzyme A). Acyl chains provided by ACP are endogenously synthesised whereas acyl chains delivered by acyl-CoA are derived from exogenous fatty acids (Weimar *et al*. 2002). 1,2-diacyl-*sn*-glycerol-3-phosphate (phosphatidic acid) is also formed in a separate pathway via phosphorylation of diacylglycerol by the kinase DgkA. Phosphatidic acid is then converted to cytidine diphosphate-diacylglycerol (CDP-diacylglycerol) by CdsA. CDP-diacylglycerol sits at the branch point in the synthesis of the major phospholipids of *E. coli*: PE, PG and CL. PE is synthesised in two steps. CDP-diacylglycerol is first converted to phosphatidylserine by PssA. PssA is located in the cytoplasm and reported to bind with high affinity to ribosomes (Louie & Dowhan 1980). However, preferential binding with the cell membrane can be induced by enrichment at the cell membrane of either CDP-diacylglycerol or phosphatidylserine. PssA is therefore thought to transition between the two sites in a substrate-dependent manner (Dowhan 1992). Phosphatidylserine is converted to PE by the decarboxylase Psd.

The other branch in the phospholipid biosynthesis pathway forms the acidic phospholipids. PG is synthesised via a two-step mechanism. CDP-diacylglycerol is first converted to phosphatidylglycerol-3-phosphate (PGP) by PgsA, and is then dephosphorylated by one of 3 phosphatidylglycerolphosphatases, PgpA, PgpB or PgpC (Funk *et al*. 1992; Icho & Raetz 1983). All three phosphatidylglycerolphosphatases are IM localised and can independently be deleted (Funk *et al*. 1992). However, if all three genes are deleted the resulting mutant is non-viable (Lu *et al*. 2011). PgpA and

PgpC are reportedly PGP-specific phosphatases, however PgpB is reported to dephosphorylate additional substrates including undecaprenyl diphosphate (Dillon *et al.* 1996; Ghachi *et al.* 2005). After separation of the IM and OM via sucrose gradient, PgpB has been identified in both fractions and reported to have dual localisation to the OM in addition to the IM (Icho 1988). The substrate activity of PgpB varies depending upon sub-cellular localisation (Icho 1988).

Finally, CL can be synthesised from the condensation of two PG molecules by the action of ClsA and ClsB synthases (Pluschke *et al.* 1978; Nishijima *et al.* 1988; Guo & Tropp 2000). Alternatively, CL can also be synthesised by the ClsC synthase, which forms cardiolipin by condensing a molecule of PG and a molecule of PE (Tan *et al.* 2012).

Modification of the cellular phospholipid levels can have dramatic and far-reaching effects. Loss of CL or PE results in altered LPS length or structure, respectively, increased blebbing, a decrease in IM tethering to the OM (most pronounced in stationary phase), an increase in the amount of unfolded OMP OmpF, and alterations in cellular glucose flux (Rowlett *et al.* 2017).

1.3.5 Cell wall biosynthesis

The cell wall is a single layered sheath of peptide-crosslinked glycan strands that encapsulate the cell in a mesh-like net. The cell wall provides a rigid structure that withstands turgor pressure and defines the overall cell shape (Huang *et al.* 2008; Höltje 1998). The *E. coli* cell wall is made up of glycan strands of repeating *N*-acetylmuramic acid (MurNAc) and *N*-acetylglucosmine (GlcNAc) units. The glycan strands are covalently crosslinked by peptide bridges attached to the MurNAc residue

(Glauner & Höltje 1990). Typically, the pentapeptide is L-Ala-D-Glu-*m*-Dap-D-Ala-D-Ala (*m*-Dap, *meso*-diaminopimelic acid; Liu & Breukink 2016; Egan *et al*. 2015). Cell wall synthesis begins in the cytoplasm, first, a UDP-MurNAc-pentapeptide molecule is synthesised from UDP-GlcNAc by the Mur enzymes MurA-F (Fig. 1.8; Marquardt *et al*. 1992; Duncan *et al*. 1990; Liger *et al*. 1995). The MurNAc-pentapeptide molecule is then transferred to a membrane anchored undecaprenyl phosphate by MraY, forming lipid I (Chung *et al*. 2013). The last step at the inner leaflet of the IM is the addition of a GlcNAc residue by the glycosyltransferase MurG to form lipid II (Fig. 1.8; Mengin-Lecreulx *et al*. 1991). The entire complex is transferred to the outer leaflet of the IM by the flippase MurJ. The disaccharide units are polymerised by glycosyltransferases to form glycan chains. The glycan chains are then crosslinked via the pentapeptide moiety by transpeptidases, to form a mesh-like cell wall structure. Typically, the most abundant type of pentapeptide crosslink is through the 3rd and 4th amino acids of two neighbouring pentapeptides, forming a 3-4 crosslink. The transpeptidase reaction results in loss of the terminal D-ala residue.

1.3.6 Transport of macromolecules across the periplasm

Proteins destined for the periplasm or the OM first need to be translocated across the IM. These proteins are directed to the heterotrimeric SecYEG complex by an N-terminal signal sequence and translocated across the IM into the periplasm where the signal sequence is then cleaved by periplasmic signal peptidases (Dalbey & Chen 2004). Unfolded OMPs are trafficked to the OM via the soluble chaperones Skp, SurA and DegP (Fig. 1.9). The protein SurA has a high-affinity for peptides with a motif

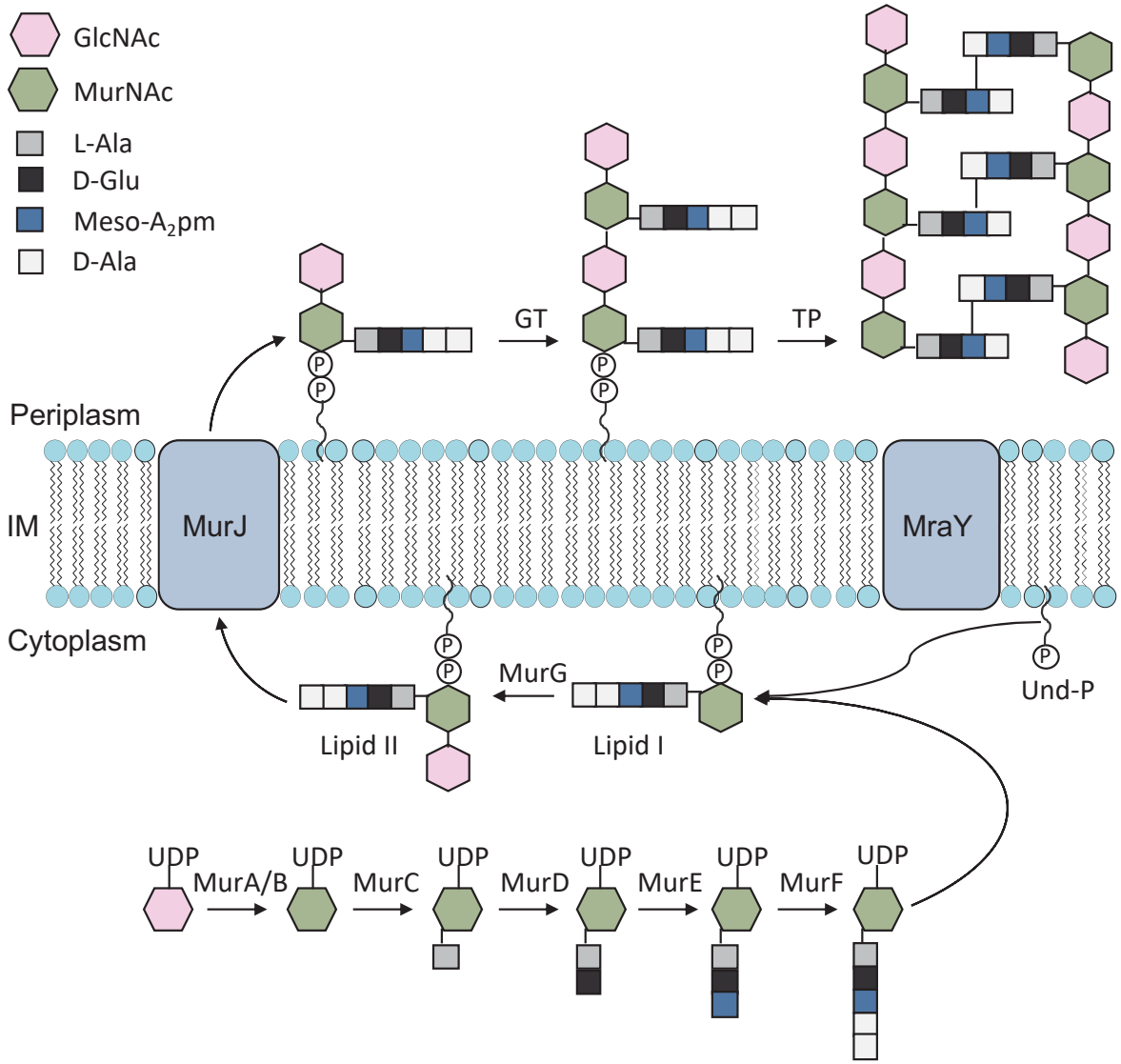

Figure 1.8 Peptidoglycan biosynthesis

A schematic representing the biosynthesis of peptidoglycan. *N*-Acetylglucosamine (GlcNAc) and *N*-Acetylmuramic acid (MurNac) disaccharides covalently linked to a pentapeptide stem form the basic repeating subunit of the cell wall, which is synthesized in the cytoplasm by the Mur enzymes. The disaccharide-pentapeptide unit is anchored to the IM by Und-PP forming lipid II. Lipid II is flipped across the IM by MurJ and repeating units are ligated by glycosyltransferases (GT) to form glycan strands, which are then crosslinked by transpeptidases (TP). (Abbreviations: L-Ala, L-Alanine; D-Glu, D-Glucose; Meso-A_2pm, meso-diaminopimelate; D-Ala, D-Alanine; UDP, uridine diphosphate; Und-P, Undecaprenyl-phosphate.)

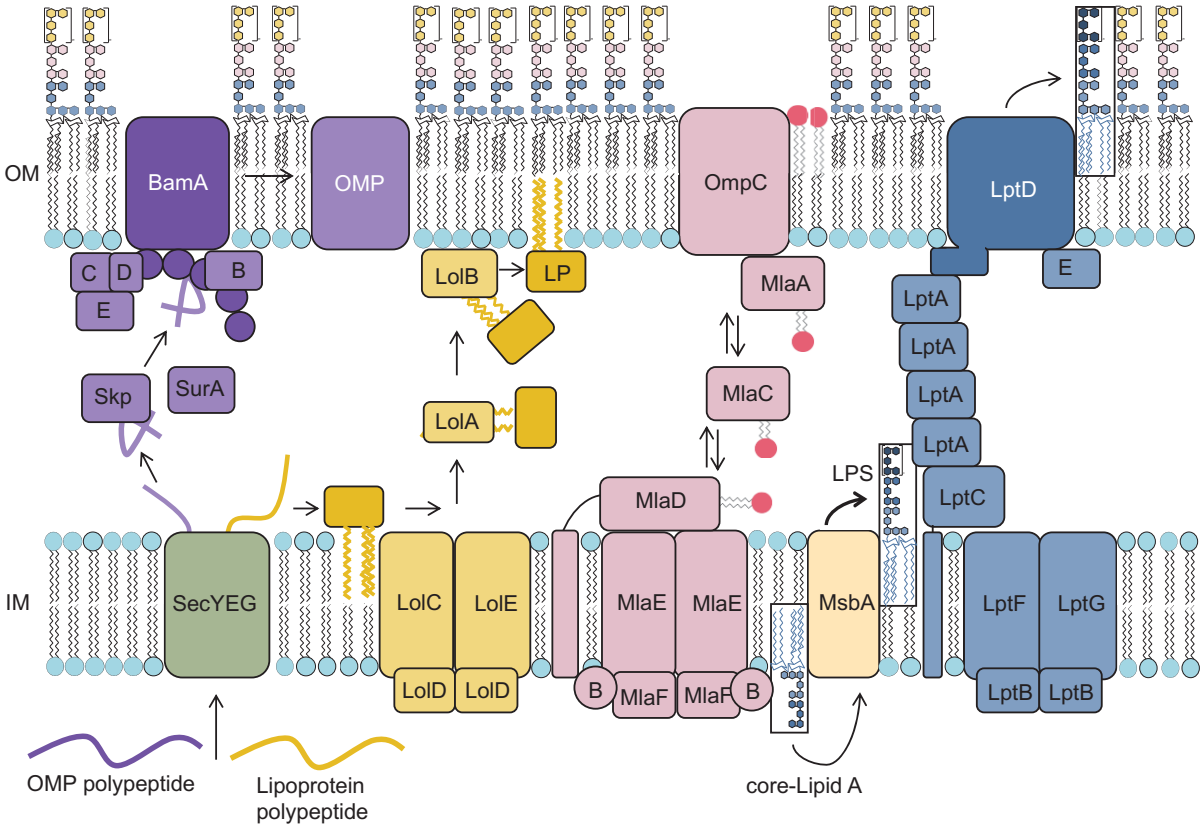

Figure 1.9 Transport of outer membrane components across the periplasm

Transport machineries are coloured according to their transported substrate: outer membrane protein (OMP) trafficking and insertion (purple); lipoprotein transport (yellow); phospholipid transport (red); lipopolysaccharide (LPS) transport (blue).

often found in OMPs, deletion of *surA* results in a reduction in OMPs and SurA has been found cross-linked to BamA *in vivo* (Bitto & McKay 2003; Sklar *et al.* 2007). SurA is therefore thought to be the main chaperone for delivery of nascent OMPs to the BAM complex. Skp has also been reported to bind denatured OMPs in the periplasm, but not denatured periplasmic proteins, and deletion of *skp* results in a decrease in specific, but not all, OMPs including OmpC and OmpF (Chen & Henning 1996). DegP, however, is reported to have both chaperone and protease activity depending upon the growth conditions. DegP is reported to aid the folding of the periplasmic protein MalS at lower temperatures, while at higher temperatures DegP is a serine protease that degrades aberrantly folded periplasmic proteins (Spiess *et al.* 1999). DegP is therefore thought to primarily function as a protease and to capture proteins that fall off the SurA/Skp pathways, especially under conditions of stress. In its native state DegP forms a trimer that is able to form larger multimeric structures, such as 12- 24- or 30-mers, which encase the proteins for degradation. Following substrate cleavage, the cage structure disassembles (Kim *et al.* 2011). Finally, double deletions of *surA* and *skp* or *surA* and *degP* are synthetically lethal, resulting in the hypothesis that the SurA and Skp pathways traffic proteins independently and as such there are two parallel pathways for the trafficking of OMPs to the BAM complex (Sklar *et al.* 2007; Rizzitello *et al.* 2001).

Lipoproteins destined for the OM are trafficked from the Sec translocon via the Lol pathway (Matsuyama *et al.* 1995). Lipoproteins secreted through the Sec translocon are first anchored into the membrane by their N-terminal signal peptide. The pre-lipoprotein is matured to a lipoprotein via several steps. First, a diacylglycerol moiety is transferred from PG to a specific N-terminal cysteine of the

pre-lipoprotein, the signal peptide is then cleaved and the diacylated cysteine residue becomes the +1 residue (Sankaran & Wu 1994; Hantke & Braun 1973). Next, the newly available N-terminal amino-hydroxide group is acylated forming a tri-acylated lipoprotein (Hantke & Braun 1973). Lipoproteins destined for the OM are distinguished from lipoproteins to be retained in the IM by the absence or presence of a Lol-avoidance signal, respectively, which is defined as an aspartate at the +2 position of the mature protein (Zückert 2014). Proteins with any other residue at the +2 position are translocated to the OM via the Lol pathway (Fig. 1.9; Yamaguchi *et al.* 1988). The Lol pathway consists of 5 proteins, LolABCDE, which are all essential. The proteins LolA and LolB were assumed to be essential because OM lipoproteins are involved in the assembly of every major OM biosynthesis complex (e.g. BamBCDE, MlaA and LptE). However recently published data suggest LolA and LolB are essential because of the resulting toxicity caused by accumulation of mislocalised lipoproteins if LolA or LolB are deleted (Grabowicz & Silhavy 2017). The proteins LolC, LolD and LolE form an ABC transporter complex in the IM that transfers IM anchored lipoproteins to LolA. LolA then cycles through the periplasm delivering lipoproteins to the OM lipoprotein LolB.

While the trafficking pathways of LPS, OMPs and OM lipoproteins are well characterized, the mechanisms of phospholipid trafficking between the two membranes has remained elusive until recently and is very much still an area of microbiology of dynamic debate. The Mla pathway, named for its role in the maintenance of OM lipid asymmetry, although not essential, is currently thought to be the primary pathway for shuttling phospholipids between the IM and OM layers (Malinverni & Silhavy 2009). The Mla pathway consists of 6 proteins, MlaABCDEF,

which between them occupy each compartment of the cell. Proteins MlaBDEF form an ABC transporter in the IM, MlaC is a periplasmic shuttle protein and MlaA is an OM anchored lipoprotein. The structure of MlaA has been solved, MlaA forms a ~20 Å donut-like structure imbedded into the inner leaflet of the OM with a central amphipathic pore for phospholipid trafficking (Abellón-Ruiz *et al.* 2017). This structure is thought to preclude phospholipid access from the inner leaflet. Instead, MlaA is thought to bind with PLs of the outer leaflet of the OM (Abellón-Ruiz *et al.* 2017). Initial research reports that this pathway acts in a retrograde direction, transporting phospholipids from the OM towards the IM (Malinverni & Silhavy 2009). However, recent preprint data from Tim Knowles' group suggest that phospholipid loading is directional from MlaD to MlaC and overall phospholipid transport is anterograde (Hughes *et al.* 2018). Thus, the direction of phospholipid trafficking of the Mla pathway remains to be confirmed, and the mechanism of anterograde phospholipid transport remains to be fully elucidated. However, in addition to MlaD, *E. coli* has 2 other MCE (mammalian cell entry) domain-containing proteins YebT and PqiB, which belong to the YebST and PqiABC systems, respectively. The proteins YebT and PqiB are reported to span the periplasm and hypothesized to transport phospholipids (Ekiert *et al.* 2017). Remarkably, deletion of all three systems does not alter the growth rate of *E. coli* in LB, although the OM surface displays an increase in ruffling (Ekiert *et al.* 2017). The viability of a Δ*mlaE* Δ*pqiA* Δ*yebS* triple deletion mutant is indicative of additional mechanisms of phospholipid trafficking in *E. coli*.

LPS is transported to the IM via the Lpt pathway. The Lpt complex spans the periplasm. Core-lipid A is first transferred across the IM by the flippase MsbA, a process which uses ATP (Zhou *et al.* 1998). Mature LPS is then extracted from the IM

and transferred to LptC, powered by the LptBFG ABC transporter. The LptBFG complex also generates the power to push LPS along the LptA bridge that joins the IM protein LptC with the OM protein LptD. LPS is finally translocated across the OM by the LptDE proteins (Simpson *et al.* 2015). The speed of LPS transport across the periplasm is estimated to be as high as ~70,000 molecules per minute (Lima *et al.* 2013).

Lastly, a strikingly absent component from the model presented in figure 1.9 is the cell wall. The mechanism(s) of action for transport of OM components across the peptidoglycan layer are largely unknown. However, structural analysis of the phospholipid trafficking systems YebT and PqiB by cryo-electron microscopy revealed sections of the structures that appear altered at the approximate position where the peptidoglycan is expected to be. It is thought that PqiB might slot directly through the cell wall, while rings 4 and 5 of YebT sandwich the peptidoglycan layer (Georgia Isom, personal communication). However, the mechanisms by which PqiB and YebT are assembled through the cell wall remain unknown. The synthesis and transfer of macromolecules through the cell wall is an even more interesting conundrum when considered alongside the reports that *E. coli* can reversibly transition between its native state and an L-form (cell-wall free) state (Errington 2017). Following transition from L-form to rod-shaped bacterium, are these periplasmic-bridging structures synthesized through the PGN once the PGN is in place, or is the cell wall synthesized around the membrane-linking structures?

1.4 Genetic screening as a tool to link phenotype and genotype

An ongoing challenge in the field of microbial genetics is to determine the function of an uncharacterised gene, a subject that is fundamentally important for a diverse range of biological questions. Certain genes may have properties that can be harnessed in the biotechnology industry. One well known example of this is the isolation of the DNA polymerase I gene from the organism *Thermus aquaticus*, which codes for the thermostable Taq polymerase, used worldwide for routine PCR (Chien *et al.* 1976; Lawyer *et al.* 1993). In industry, it is important to understand which genes contribute to fitness in order to streamline the organism for most efficient growth. In recent decades, there has been an increase in the drive to characterise genes that contribute to virulence of a pathogen, or that confer resistance to antibiotic treatment. Improved understanding of the molecular mechanisms underlying virulence, antibiotic resistance or antibiotic targets will enable better treatment options.

E. coli has been described as the 'workhorse of molecular microbiology' due to its extensive use in laboratories to understand gene function (Blount 2015). As such there is a wealth of characterisation data available, but even for the model organism *E. coli* there remains a number of genes of unknown function. One report states this number is as large as a third of the genome, although this figure is likely lower given the time span since estimation (McAteer *et al.* 2001). More recent studies suggest around 80% of genes have an annotated function, but only 54% have experimental evidence to support these annotations (Hanson *et al.* 2009; Frishman 2007).

There are two approaches to linking a phenotype (an observable trait) with a genotype (the genetic material that gives rise to a phenotype): forward and reverse genetics. Forward genetics involves the construction or isolation of a mutant with a

phenotype, and then identification of the underlying mutation that causes the phenotype. While reverse genetics involves identification of the phenotypic effects of a given mutation. The origins of forward genetics stemmed from the characterisation of auxotrophic strains that could not synthesise specific amino acids by Joshua Lederberg (Lederberg 1946).

1.4.1 A History of genetic screening

Historically, genetic studies to map gene function relied on X-ray or UV mutagenesis to create mutants. Isolated mutants were further characterised to identify the mutation site(s) and a related phenotype. Early work was centred on genes involved in the biosynthesis of amino acids (Lederberg & Tatum 1946). Lederberg isolated two mutants of *E. coli* following X-ray treatment that could not synthesise methionine or proline, respectively, and therefore could only grow if the medium was supplemented with the amino acid they were unable to synthesise. In collaboration with Edward Tatum, he extended his assay to use two double mutants in an *E. coli* K-12 background that could not synthesise methionine and biotin, or threonine and proline and discovered that when the two mutants were mixed together it was possible to isolate new mutants able to synthesise both methionine and proline, concluding that genetic material must have been shared. Lederberg later described this process as conjugation.

It was then realised that during conjugation the transfer of genetic material from Hfr F^+ strains was always initiated from the same locus (*oriT*), that the rate of transfer was consistent, and that donor alleles were transferred in the same order. This knowledge led to the development of interrupted mating, whereby two strains

undergoing conjugation were interrupted after a given time point and the extent of material transferred could be deduced from phenotypic studies. By measuring the length of time it took to transfer the required material from one strain to another, the heritable unit that conferred a trait could be mapped to an approximate position around the chromosome (Wollman & Jacob 1955). In turn, the relative positions of characterised genes were identified and given in minutes to produce a map of the *E. coli* chromosome (Taylor & Thoman 1964; Taylor & Dunham Trottfer 1967).

However, it wasn't until 1965 that the first sequencing technologies were devised, initially applied to tRNA and rRNA molecules (Holley *et al.* 1965; Sanger *et al.* 1965). In 1972 the first complete coding sequence for a protein, the bacteriophage MS2 coat protein, was determined, and in 1977 the first complete DNA genome was sequenced: the genome of the bacteriophage ϕX174 ('PhiX'), which is used today as a positive control in Illumina sequencing technologies (Min Jou *et al.* 1972; Sanger *et al.* 1977). With the use of mutagenesis, linkage mapping, and Sanger sequencing it became possible to comprehensively isolate the underlying genetic cause of a phenotype. The simultaneous progression in microbial genetics was the ability to construct stable mutants by design, rather than relying on the random mutagenesis introduced by UV light, X-rays or chemicals.

In time, the number of characterised genes expanded. A challenge in assigning function to remaining uncharacterised genes can be the difficulty in identifying a phenotype. Certain genes might have a very limited niche-specific role, lowering the chances of identifying a phenotype when screening a single mutant under one condition at a time. A means to combat this was the development of high-throughput screens to allow screening of several mutants and/or conditions at once. An early

example of this is the use of replica plating to transfer cells from a master plate with several hundred mutant colonies on up to 20 different replica plates, thereby enabling the screening of thousands of mutants in up to 20 conditions in under an hour (Hasunuma 2009).

Modern day high-throughput screens involve the construction of mutant libraries. Two general mechanisms exist for construction of large collections of mutants: random mutagenesis and targeted mutagenesis. A strength of using targeted mutagenesis, such as the construction of gene-deletion mutants, is that the mutants can be individually archived to create an ordered library. Unlike a library of pooled mutants, these mutants are easily accessible and can be used for reverse genetics screening too. Ordered storage of transposon mutants is possible but it can require extensive sequencing capabilities and robotics (Goodman *et al.* 2009). However, a strength of transposon mutagenesis is that experiments are not limited by the annotation quality of the genome and the full breadth of the genome can be assayed, including intragenic regions that are often overlooked.

One of the most well-known ordered libraries is the Keio collection of gene-deletion mutants (Baba *et al.* 2006). This library was used to study the essential genes of *E. coli* K-12, identified as those that could not viably be deleted. Copies of this library are commercially available and have been used for a wide range of biological screens, including stress and antibiotic assays to evaluate mutant fitness under a diverse range of conditions, cell permeability assays and genetic linkage mapping (Nichols *et al.* 2011; Liu *et al.* 2010; Typas *et al.* 2008).

The use of transposon mutagenesis increased tremendously with the coupling of 'next generation' sequencing (NGS) to identify mutation sites. Rapid identification

of the transposon insertion site increases the number of mutants that can easily be profiled in a single experiment. While the Keio library permits profiling of ~4,000 mutants, transposon mutagenesis can enable profiling of >1,000,000 mutants. With the arrival of NGS technologies and increase in library complexity, a whole genome screen of the phenotypic profile of a compound can be determined at the codon level of resolution (Goodall & Robinson *et al.* 2018).

1.4.2 History of the transposon

A transposon, or transposable element, is defined as a DNA sequence that can move location within a genome; colloquially given the name "jumping gene". The transposon was first described by Barbara McClintock, reportedly as early as 1948. Her work, published in 1950, described the ability of a linear fragment of DNA to insert at different positions on the maize chromosome and reversibly inactivate the target gene into which it inserted (McClintock 1950; Ravindran 2012). McClintock had discovered two independent genetic loci that she named the "*Dissociator*" and the "*Activator*" and together named them "controlling elements". She found that the *Activator* (*Ac*) controlled transposition of the *Dissociator*, (*Ds*) and that the transposition of *Ds* involved cleavage of the genome. The *Ac* and *Ds* were later confirmed to be type II transposons. The *Ac* and *Ds* transposons were shown to be structurally related, but the *Ds* transposon was lacking a functional transposase and unable to transpose without exogenous transposase, congruent with McClintock's finding that transposition of *Ds* is dependent on *Ac*. However, at the time of discovery her work was reportedly met with uncertainty and doubt. The concept of jumping genes, although ground-breaking, was far removed from the understanding of

genetic material at that period in time. By the end of the 1960's there was increasing evidence for mobile genetic elements in bacteria and bacteriophage and the importance and recognition of her discovery was later realised (Fedoroff 1995).

In the 1970's two groups independently identified the ability of Tn*5* and Tn*10* transposons to transpose antibiotic-resistance genes. The Tn*10* transposon was characterised and well-studied by Nancy Kleckner who first showed that the Tn*10* transposon conferred resistance to tetracycline (Kleckner *et al.* 1975). The Tn*5* transposon was discovered fortuitously by Douglas Berg while studying the transfer of kanamycin resistance from R factor to λ phage (Berg *et al.* 1975). The Tn*5* transposon actually encodes three genes that confer resistance to kanamycin, bleomycin and streptomycin. In *E. coli*, however, the streptomycin resistance gene is cryptic (Auerswald *et al.* 1981; Mazodier *et al.* 1985). The antibiotic resistance genes are sandwiched between two almost identical inverted sequences (IS) termed IS*50*L and IS*50*R. IS*50*L and IS*50*R differ in sequence by one nucleotide (Mazodier *et al.* 1985). The IS*50*R sequence codes for two proteins translated from a single transcript: a transposase and a protein that inhibits transposition (Johnson *et al.* 1982; Isberg *et al.* 1982). The point mutation in IS*50*L results in a premature stop codon and a truncated, non-functional transposase. However, the point mutation also increases expression of the kanamycin resistance gene (Rothstein *et al.* 1980; Rothstein & Reznikoff 1981).

The Tn*5* transposon is a class II transposon; transposition occurs via a "cut and paste" mechanism. The transposon is excised precisely from its original position, but inserts into a new position via a staggered cut in the target DNA, resulting in a 9 bp duplication of target DNA either side of the inserted transposon (Berg *et al.* 1983).

The transposition reaction is catalysed by a transposase enzyme encoded by the *tnp* gene within the Tn*5* sequence, however, transposase can also be provided exogenously (Reznikoff 1993). The transposase recognises a 19 bp inverted repeat sequence at the terminal ends of the transposon sequence. The use of this specific 19 bp repeat is employed in the mini-Tn*5* constructs used today for construction of dense transposon mutant libraries.

1.4.3 Transposon mutagenesis

The ability to deactivate a target gene by insertion of a transposon is a useful tool in bacterial genetics. In the absence of a transposase, the resulting mutant is often stable and therefore suitable for phenotypic studies. The transposon can carry a selectable marker enabling easy and rapid identification of successful transformants. The transposon can also carry gene constructs that will create fusion proteins when inserted into a coding sequence. These constructs can then be used for phenotypic screening, such as the identification of secreted or membrane proteins using TnPhoA (de Lorenzo *et al.* 1990; Manoil & Beckwith 1985). Tn*5* is particularly useful as a genetic tool as it transposes into a wide-range of Gram-negative bacteria at a high frequency and has low/no target specificity. As such, it has been used widely in a number of species (de Bruijn 1987; de Bruijn & Lupski 1984; Beringer *et al.* 1978).

The use of transposons for whole-genome phenotypic screening in a single experiment was first reported in *Saccharomyces cerevisiae* and given the name "genetic footprinting" (Smith *et al.* 1995; Smith *et al.* 1996). The purpose of the experiment was to identify phenotypes for genes of unknown function and thereby assign functions to newly annotated orphan genes. The assay used a Ty*1* transposable

element to create a mutant population. Representative samples of the mutant population were independently exposed to different conditions, and the Ty*1* insertion sites compared between the input and output pools of mutants. The Ty*1* insertion sites were determined by PCR amplification of the transposon junction. A genomic DNA primer, designed specifically for the gene of interest, introduced a fluorescent probe to the PCR products, which were then separated by gel electrophoresis. The transposon insertion site was estimated from the length of the separated fragments. If the target gene was required for growth under the condition of interest, transposon-mutants were lost from the output population. This was identified by an observable difference in PCR-fragment patterns when separated in an agarose gel, hence the name 'footprint'. The function of a gene was then inferred from the condition-dependent footprint.

Genetic footprinting was subsequently adapted for use in several bacterial species, including *E. coli* (Hare *et al*. 2001; Badarinarayana *et al*. 2001), *P. aeruginosa* (Wong & Mekalanos 2000), *Helicobacter pylori* (Jenks *et al*. 2001), *Mycoplasma genitalium* (Hutchison *et al*. 1999) *Haemophilus influenzae* (Reich *et al*. 1999; Akerley *et al*. 2002) and *Streptococcus pneumoniae* (Akerley *et al*. 1998).

In 2001, Hare *et al*., using a mini-Tn*10* transposon, optimised the technique by incorporating an outward facing promotor into the transposon to reduce polar effects, and an *in vivo*, plasmid-borne transposition method involving a plasmid containing a transposase linked to a regulatable promoter to prevent further mutagenesis (Hare *et al*. 2001).

In these early studies the transposon insertion sites were identified by PCR-based mapping. While this method is accurate, the process is laborious as transposon

insertions are detected individually using position-specific PCR primers, which is an unrealistic method for a high-throughput whole genome screen. Transposon mutagenesis was subsequently coupled with the use of DNA microarrays to allow pooled transposon-junction analysis. Two different methods arose: signature-tagged mutagenesis (STM), and transposon site hybridisation (TraSH) (Hensel *et al.* 1995; Mei *et al.* 1997; Sassetti *et al.* 2001; Sassetti *et al.* 2003; Brown & Botstein 1999; Badarinarayana *et al.* 2001). STM makes use of individually barcoded mutants allowing specific mutant identification post screening. However, STM is limited to comparatively smaller pool sizes than TraSH and is somewhat labour intensive. While the method was more high-throughput than PCR-based mapping, issues were reported with a high false-positive discovery rate and poor correlation between independent microarrays (Tong *et al.* 2004).

1.4.4 Transposon sequencing

With the development of short-read sequencing technologies for whole-genome sequencing, as many as 24 million short DNA fragments could be sequenced in under 24 h in a single MiSeq run. The Illumina sequencing platform was adapted for sequencing the transposon junctions of pooled transposon mutants. In 2009 four independent research groups published transposon insertion sequencing methods: HITS, TraDIS, INSeq and Tn-seq (Gawronski *et al.* 2009; Langridge *et al.* 2009; Goodman *et al.* 2009; van Opijnen *et al.* 2009). All four protocols share the same overarching protocol. Collectively these studies consist of three steps: random transposon mutagenesis to create a mutant population, selective outgrowth of a sample population, and sequencing of the transposon junction to detect the

transposon insertion location. Despite their similarity, the different protocols vary in the precise method, transposon used, and research question. The first TraDIS (Transposon Directed Insertion-site Sequencing) and Tn-seq experiments were used to determine gene essentiality in rich medium and the fitness effect of disruptions, while INSeq and HITS were used to test human infection models. INSeq and Tn-seq use a *Mariner* transposon, making use of a single nucleotide change to introduce an *Mme*I type IIS restriction site within the transposon. The restriction enzyme makes a staggered cut 20 bp downstream from the recognition site in the transposon, 16 bp into the genomic DNA, which is sufficient to sequence and identify the insertion location.

The HITS and TraDIS methods lack type IIS restriction steps, instead, fragments are generated through shearing of genomic DNA. The HITS protocol includes an additional purification step to remove contaminating genomic DNA before sequencing. The shearing method then requires additional processing to ligate adaptors in preparation for sequencing, as well as variation in DNA fragment size which may have implications for PCR bias during preparation of sequencing libraries. However, an advantage of these two methods is that they can be applied to a wide range of transposons and insertion elements, and are especially more effective in strains with a higher GC content as the mariner transposon inserts specifically into TA sites (van Opijnen & Camilli 2013; Goodman *et al*. 2009; Bryan *et al*. 1990). In short, the optimal protocol to use is dependent upon the strain, the research question and the lab resources available.

1.4.5 Applications of transposon sequencing

The method chosen for the work in this thesis is TraDIS. TraDIS is a negative selection assay that simultaneously screens thousands of transposon mutants and maps the location of the transposon insertion point using next generation sequencing technologies, first published by Langridge *et al.* (2009). The high number of mutants generated enables full-genome screening in one experiment, allowing the simultaneous assay of every gene.

An overview of the TraDIS method is presented in figure 1.10. The first step in the protocol is construction of a transposon library. Typically, in *E. coli*, a mini-Tn*5* is transformed by electro-transformation. In theory, each bacterial cell has a single transposon inserted at random into its genome. The mini-Tn*5* transposon is an ideal choice because it has low/no reported insertion bias and is small in size, increasing the transformation efficiency. Successful transformants are then selected for on supplemented medium. The resulting colonies are pooled to form a library of transposon mutants. To identify the transposon insertion site, genomic DNA is extracted and the gDNA-transposon junction is amplified and sequenced. The sequence data are mapped to a reference genome and can be used to identify the genome-wide insertion sites. Together, these data can be used to identify essential and non-essential genes of the genome. Generally, a gene that cannot be viably disrupted by a transposon is assumed to be an essential gene, while genes that can be disrupted by transposon insertion along the full length of the coding sequence are considered to be non-essential (Fig. 1.10). However, there are some examples were a gene can have both an essential and a dispensable domain (Christen *et al.* 2014).

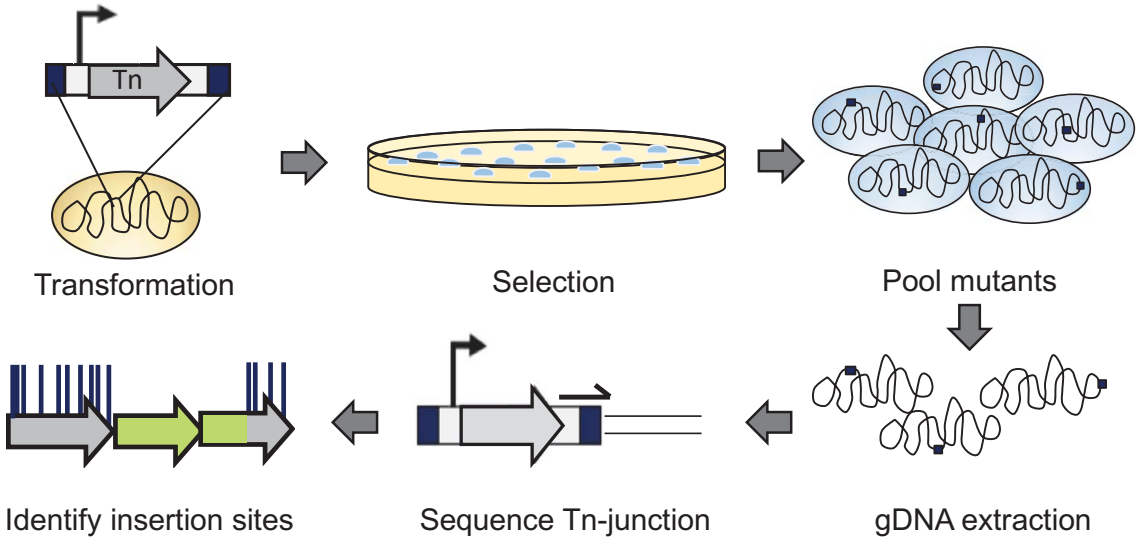

Figure 1.10 Overview of a TraDIS experiment

The sequential steps of a general TraDIS experiment, starting with transformation of a single transposon per cell and selection of successful transformants on an agar plate. Mutants are pooled and genomic DNA (gDNA) extracted. The transposon insertion site is identified by sequencing of the transposon junction. Insertion data are mapped to a reference genome, and presence or absence of insertion sites can be used to identify non-essential genes (grey), essential genes (green) and genes with essential regions, respectively.

Transposon-sequencing has been applied to a wide-range of bacteria species (Langridge *et al*. 2009; Byrne *et al*. 2014; Dembek *et al*. 2015; Phan *et al*. 2015; Subashchandrabose *et al*. 2016; Christen *et al*. 2014; Hassan *et al*. 2016; Paulsen *et al*. 2017; Grant *et al*. 2016; Troy *et al*. 2016). Overall, the technique is generally used for three purposes, (i) to identify the essential genome of an organism (ii) to identify conditionally-essential genes such as those required to survive in a defined medium, under conditions of stress or required for colonisation of a specific niche such as invasion of host tissue in an infection model (iii) or to interrogate biosynthetic pathways.

The detection of essential genes has been outlined above. Identification of conditionally essential genes is an extension of this method, but involves further growth of the transposon library under a condition of interest and comparison of the input and output pools of mutants. If a gene required for survival under a given condition is disrupted, the resulting mutant will die or be outcompeted, resulting in loss of that mutant from the population. Sequencing of the transposon junction allows quantitative analysis of both the location of the transposon and relative frequency of mutants of the same genotype within the population. These data can therefore be used to identify genes that are advantageous and disadvantageous for growth under a given condition, as well as conditionally essential genes (Fig. 1.11). This overarching method can be applied to a range of conditions, from antibiotic stress, or the ability to grow and survive in serum and urine, to the ability to colonise host tissues. Transposon sequencing can even be used to identify conditionally essential genetic requirements among a community of species occupying a niche. For example, a transposon library constructed in *Bacteroides thetaiotaomicron* (a human gut

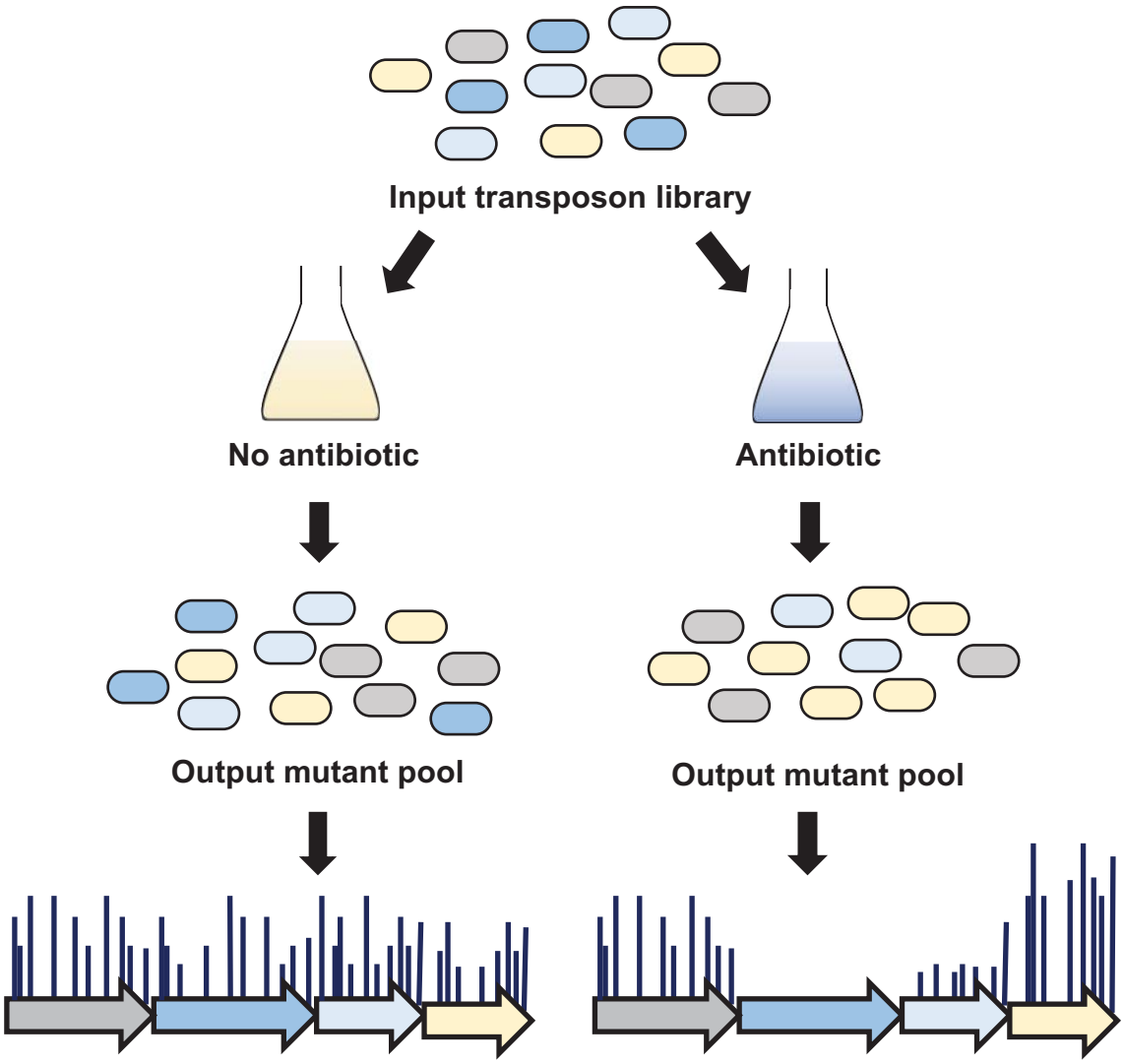

Figure 1.11 Identification of conditionally essential genes using TraDIS

Growth of a transposon library in media with and without an antibiotic. Selection pressure results in loss of some mutants from the population (blue) and enrichment of others (yellow). Addition of an antibiotic has a neutral effect on some mutants (grey). Insertion frequency can be used as a measure of mutant abundance. Comparison of the input and output pools between conditions can identify conditional essential genes (dark blue), genes that are advantageous for growth in the presence of the antibiotic (pale blue) and genes that, when disrupted, confer a selective advantage under the condition of interest (yellow).

commensal) was used to colonise immunodeficient mice in the presence or absence of additional bacterial species. An operon in *B. thetaiotaomicron* is regulated by vitamin B_{12} and was conditionally essential for the colonisation of the mouse gut in the absence of any other bacterial species, but in the presence of Firmicute and Actinobacteria, the requirement for this operon is alleviated (Goodman *et al.* 2009). As *B. thetaiotaomicron* cannot synthesise its own vitamin B_{12}, this result revealed the operon involved in vitamin B_{12} uptake during host colonisation.

The third use of transposon mutagenesis is the identification of synthetic lethal partners, which can be used to map biosynthetic pathways. The term 'synthetically lethal' is used to describe the relationship between two (or more) non-essential genes that can independently be deleted, but where the combined deletion results in a non-viable mutant. Synthetically lethal relationships between genes have been used to map genetic interactions in defined mutants, such as a study in yeast that deleted extensive pair-wise combinations of genes to identify genes involved in DNA replication and repair, and chromatin structure maintenance (Pan *et al.* 2006). In the context of transposon sequencing, it is possible to identify all genes that are synthetically lethal with your target gene of interest by constructing a transposon mutant library in a gene deletion strain (Fig. 1.12). Comparisons of insertion mutants between a transposon library in the parent strain and the gene-deletion strain enable identification of core essential genes, non-essential genes, synthetically lethal genes and revertant non-essential genes (genes that are no longer essential in the gene-deletion transposon mutant library).

It is also possible to identify synthetically lethal relationships without construction of the gene-deletion mutant via the use of chemicals that inactivate the

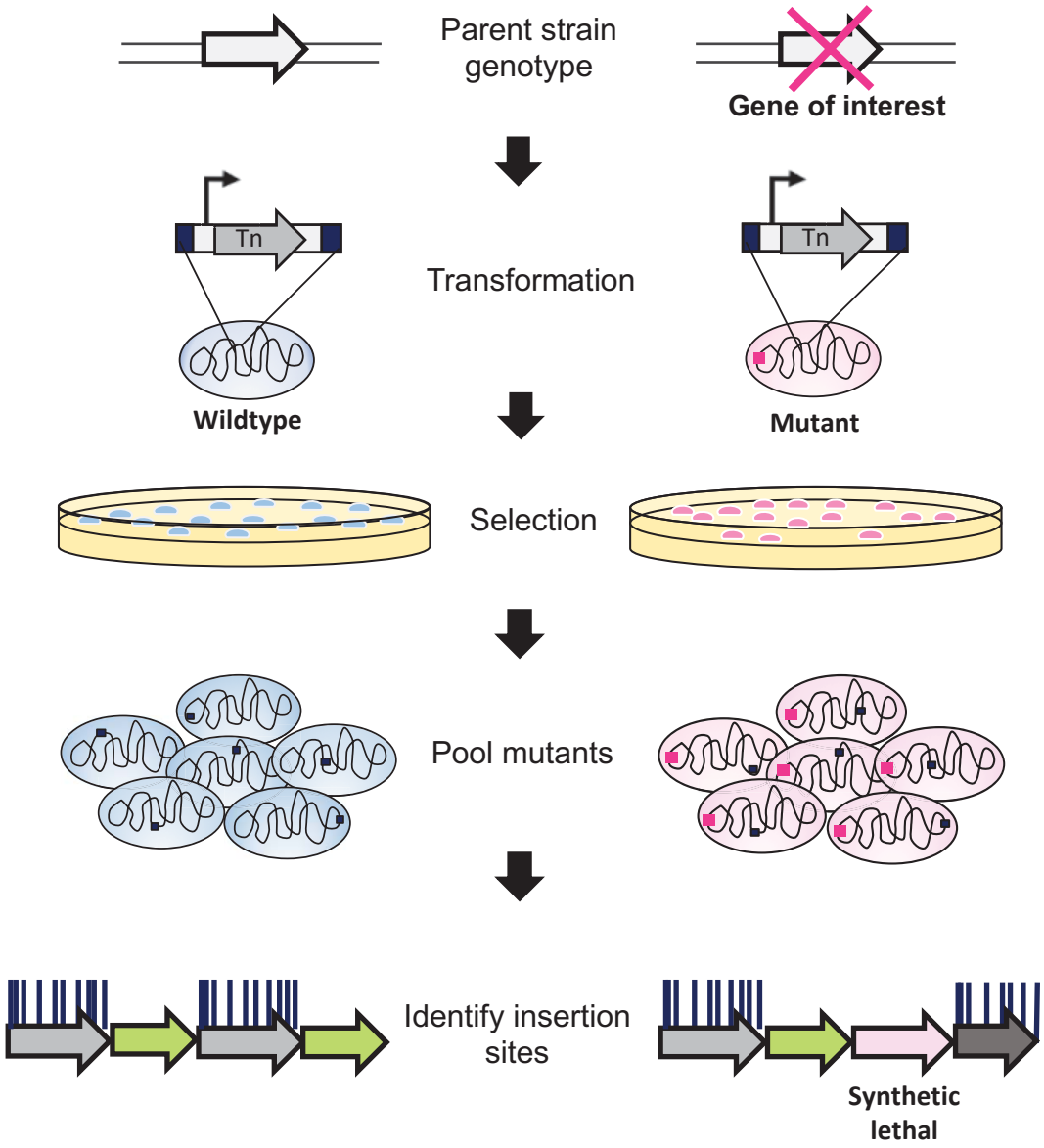

Fig. 1.12 Identification of synthetically lethal gene pairs using TraDIS

Comparison of transposon mutant libraries constructed in an engineered mutant (pink) and its parent strain (blue) reveals core essential genes (green) that cannot be disrupted and dispensable genes (grey). A gene that cannot be disrupted in the mutant background but is dispensable in the wildtype strain is synthetically lethal with the deleted gene of the mutant (pink). Conversely, a gene that is essential in the parent strain, but can be disrupted in the engineered mutant is a conditionally dispensable gene (dark grey).

gene or pathway of interest. An example of this is the use of the small molecule inhibitor, tunicamycin, which targets the first step in wall teichoic acid synthesis in *S. aureus*. *S. aureus* has two types of teichoic acids that make up the outer layer of the cell together with peptidoglycan. Both teichoic acids: lipoteichoic acids and wall teichoic acids, are critical for maintaining envelope integrity. A transposon library constructed in *S. aureus* was treated with tunicamycin to identify genes that have a synthetically lethal relationship with the synthesis of wall teichoic acid (Santa Maria *et al*. 2014). This approach identified the requirement for D-alanylation of lipoteichoic acids in the absence of wall teichoic acids. Transposon sequencing can therefore be used to infer genetic interaction networks, a finding that was also recognised by Opijnen *et al.* (2009) in their initial publication of Tn-Seq. Opijnen *et al*. (2009) constructed a transposon library in *S. pneumoniae*, as well as 5 transposon libraries in single gene-deletion mutants: 3 transcriptional regulator genes and 2 carbohydrate-uptake genes. They used this data to infer a wide genetic interaction network for the regulation of carbohydrate uptake.

1.5 Aims

The first aim of this thesis is to use transposon sequencing of a very dense transposon library to identify the essential genome of the model organism *E. coli* K-12 and benchmark these data against the gold-standard Keio collection of deletion mutants. Given the density of this library, it is also a powerful tool to tease apart the intricacies of transposon mutagenesis on a whole genome scale. The second aim of this thesis is to extend the TraDIS method to identify the conditionally essential genes of *E. coli* required for growth in the presence of the clinically relevant antibiotic polymyxin B. The final aim was to validate the approach by further characterising any genes of unknown function that may yield promising leads as novel antimicrobial targets.

Chapter 2

Materials and Methods

2.1 Bacterial growth media

2.1.1 Culture media

All media were made using distilled water and sterilised by autoclaving at 121°C for 15 min. For standard growth experiments strains were inoculated into lysogeny broth (LB) comprised of 5 g/L NaCl, 10 g/L tryptone and 5 g/L yeast extract. 2x TY broth was used for growth of competent cells during construction of a transposon library, which consisted of: 5 g/L NaCl, 16 g/L tryptone and 10 g/L yeast extract. Super Optimal broth with Catabolite repression (SOC) medium was used in the recovery step of competent cell transformations to maximise transformation efficiency. Ready-made SOC solution was purchased from Sigma Aldrich. For growth on solid phase media, 1.5% (w/v) nutrient agar was added to LB.

For growth using a defined minimal medium, supplemented M9 media was used. The M9 medium base was made from a 5x stock (Sigma Aldrich) diluted in water and autoclave sterilised. This M9 medium was routinely supplemented with 1 ml 20% (w/v) filter sterilised casamino acids (Becton Dickinson), 200 µl filter sterilised 1 M $MgSO_4$, 10 µl filter sterilised 1 M $CaCl_2$, and 2 ml of a 20% carbon source for each 100 ml. Carbon sources included 20% (w/v) D-glucose, 20% (v/v) glycerol, 20% (w/v) acetate and 20% (w/v) succinate.

2.1.2 Antibiotic supplements

When required, media was supplemented with 50 µg/ml kanamycin sulfate (Sigma Aldrich), 100 µg/ml carbenicillin disodium salt, 35 µg/ml chloramphenicol or 15 µg/ml tetracycline hydrochloride. All stock solutions of antibiotics and other supplements were filter sterilised through 0.22 µm syringe filters and stored at -20°C unless

otherwise specified by the manufacturer. For a full list of antibiotic stock solutions and respective solvents see Table 2.1.

2.2 Bacterial strains, plasmids and growth conditions

2.2.1 Bacterial strains

Bacterial strains used in this study are outlined in Table 2.2. For long term storage, strains were stored as glycerol stocks in LB medium supplemented with 15% glycerol at -80°C. When required, strains were streaked to single colonies onto the appropriate agar plate and incubated overnight at 37°C. *E. coli* K-12 strain BW25113 is the parent strain for the Keio collection of deletion mutants. *E. coli* BW25113 was used for all subsequent experiments and mutant constructs unless otherwise stated and has the genotype: $lacI^q$ $rrnB_{T14}$ Δ$lacZ_{WJ16}$ *hsdR*514 Δ$araBAD_{AH33}$ Δ$rhaBAD_{LD78}$ (Datsenko & Wanner 2000).

2.2.2 Bacterial growth conditions

Overnight cultures were grown in 5 ml volume of LB broth in a 25 ml universal bottle and inoculated with a single colony. All other liquid cultures were grown in Erlenmeyer flasks. Liquid cultures were incubated in a shaking incubator at 180 rpm and 37°C. Growth rates were monitored by measuring optical density at 600 nm (OD_{600}) using a FLUOstar Optima (BMG Labtech, Offenberg, Germany). For a standard growth curve, 50 ml LB in a glass 250 ml Erlenmeyer flask was inoculated from an overnight culture at a starting OD_{600} = 0.05, after which the optical density was recorded hourly for 8 h.

Table 2.1 Stocks of antibiotics and antimicrobial agents used in this study

Antibiotic	Stock concentration	Solvent	Storage	Supplier
Kanamycin sulfate	50 mg/ml	Water	-20°C	Sigma Aldrich
Chloramphenicol	35 mg/ml	Ethanol	-20°C	Merck
Carbenicillin disodium salt	100 mg/ml	Water	-20°C	Sigma Aldrich
Tetracycline hydrochloride	15 mg/ml	Methanol	-20°C	Sigma Aldrich
Polymyxin B	1 mg/ml	Water	NA	Sigma Aldrich
Vancomycin hydrochloride	200 mg/ml	Water	-20°C	Cayman Chemical Company
SDS	20%	Water	30°C	Fisher Bioreagents
EDTA	250 mM	Water	RT	Sigma Aldrich

Table 2.2 Strains used in this study

Strain	Description	Source
BW25113	Parent strain of the Keio collection. Genotype: *lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR*514 Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$	Datsenko and Wanner (2000)
DH5α	Cloning strain. Genotype: *fhuA*2 (*argF-lacZ*)U169 *phoA glnV*44 80 (*lacZ*)M15 *gyrA*96 *recA*1 *relA*1 *endA*1 *thi*-1 *hsdR*17	New England BioLabs
BW25113 *guaA*::*aph*	*guaA* gene replaced by kanamycin resistance cassette	This study
BW25113 *rpmF*::*aph*	*rpmF* gene replaced by kanamycin resistance cassette	This study
BW25113 *rpmI*::*aph*	*rpmI* gene replaced by kanamycin resistance cassette	This study
BW25113 *ybbD*::*aph*	*ybbD* gene replaced by kanamycin resistance cassette	This study
BW25113 *ydaE*::*aph*	*ydaE* gene replaced by kanamycin resistance cassette	This study
BW25113 *ydcD*::*aph*	*ydcD* gene replaced by kanamycin resistance cassette	This study
BW25113 *yddK*::*aph*	*yddK* gene replaced by kanamycin resistance cassette	This study
BW25113 *ydhL*::*aph*	*ydhL* gene replaced by kanamycin resistance cassette	This study
BW25113 *yedM*::*aph*	*yedM* gene replaced by kanamycin resistance cassette	This study
BW25113 *ykiB*::*aph*	*ykiB* gene replaced by kanamycin resistance cassette	This study
BW25113 *ymfE*::*aph*	*ymfE* gene replaced by kanamycin resistance cassette	This study
BW25113 *yncH*::*aph*	*yncH* gene replaced by kanamycin resistance cassette	This study
BW25113 *ypjC*::*aph*	*ypjC* gene replaced by kanamycin resistance cassette	This study
BW25113 *ybgT*::*aph*	*ybgT* gene replaced by kanamycin resistance cassette	This study
BW25113 *rfaP*::*aph*	*ypjC* gene replaced by kanamycin resistance cassette	This study
BW25113 *galU*::*aph*	*ypjC* gene replaced by kanamycin resistance cassette	Courtesy of Rhys Humphries
BW25113 *bamB*::*aph*	*bamB* gene replaced by kanamycin resistance cassette	Courtesy of Kara Staunton
BW25113 *surA*::*aph*	*surA* gene replaced by kanamycin resistance cassette	Courtesy of Karl Dunne
BW25113 *yhcB*::*aph*	*yhcB* gene replaced by kanamycin resistance cassette	This study
BW25113 Δ*yhcB*	*yhcB* gene deletion by recombination of the kanamycin cassette FRT sites, leaving a 102 nt scar	This study

2.2.3 96-well plate growth kinetics

For a routine growth curve in a microtiter plate, Greiner bio-one 96-well U-bottom plates were used with a CLARIOostar 96-well plate reader (BMG Labtech). 200 μl media was inoculated at a starting OD_{600} = 0.02. Plates were incubated at 37°C with shaking prior to every reading; optical density was recorded every 5 min for 18 h.

2.2.4 Plasmids

Plasmids used in this study are listed in Table 2.3. When growing strains carrying a plasmid, growth media was supplemented with the required antibiotic.

2.3 Molecular genetics techniques

Unless otherwise stated, kits were used following the manufacturer's instructions.

2.3.1 preparation of genomic DNA and plasmid DNA

Genomic DNA was isolated using the STRATEC RTP Bacteria DNA Mini kit, following the protocol for Gram-negative bacteria. Plasmid DNA was isolated from an overnight culture using the Qiagen QIAprep Spin Miniprep kit. Large plasmids, such as derivatives of pRW50 were extracted using either the Qiagen QIAprep Spin Maxiprep kit, or a custom phenol/chloroform extraction method and ethanol precipitation (see below).

2.3.2 Phenol/chloroform and Ethanol extraction of DNA

DNA samples were mixed with an equal volume of Phenol/Chloroform/Isoamyl Alcohol (25/24/1 v/v, with 10 mM Tris pH 8.0; 1 mM EDTA; Sigma Aldrich) and

Table 2.3 Plasmids used in this study

Plasmid	Description	Source
pBAD-myc-His	Arabinose inducible promoter upstream of a myc tag and poly-His tag.	Invitrogen
pCP20	Temperature sensitive plasmid that expresses FLP recombinase. Amp^R	Datsenko and Wanner (2000); Cherepanov and Wackernagel (1995)
pKD4	Template for kanamycin resistance cassette flanked by FLP sites. Kan^R	Datsenko and Wanner (2000)
pKD46	Temperature sensitive plasmid with an arabinose inducible promoter. Expresses λ Red recombination proteins. Amp^R	Datsenko and Wanner (2000)
pROD62	Derivative of pROD10 expressing the mCherry gene. Amp^R	Reyes-Lamothe *et al.* (2010)
pRW224	Transcription assay vector. Tet^R	Lodge *et al.* (1992)
pRW225	Translation assay vector. Tet^R	Lodge *et al.* (1992)
pRW224-Tn1.1::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 1, ORF 1 cloned into MCS	This study
pRW224-Tn1.2::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 1, ORF 2 cloned into MCS	This study
pRW224-Tn1.3::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 1, ORF 3 cloned into MCS	This study
pRW224-Tn2.1::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 2, ORF 1 cloned into MCS	This study
pRW224-Tn2.2::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 2, ORF 2 cloned into MCS	This study
pRW224-Tn2.3::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 2, ORF 3 cloned into MCS	This study
pRW225-Tn1.1::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 1, ORF 1 cloned into MCS	This study
pRW225-Tn1.2::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 1, ORF 2 cloned into MCS	This study
pRW225-Tn1.3::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 1, ORF 3 cloned into MCS	This study
pRW225-Tn2.1::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 2, ORF 1 cloned into MCS	This study
pRW225-Tn2.2::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 2, ORF 2 cloned into MCS	This study
pRW225-Tn2.3::*lacZ*	Chloramphenicol transposon–*lacZ* fusion Tn orientation 2, ORF 3 cloned into MCS	This study
pRW224-*accB*-UTR-SC1::*lacZ*	Promoter fragment cloned into MCS	This study
pRW224-*accB*-UTR-SC2::*lacZ*	Promoter fragment cloned into MCS	This study
pRW224-accB-UTR-SC3::*lacZ*	Promoter fragment cloned into MCS	This study
pRW224-*accB*-UTR-SC4::*lacZ*	Promoter fragment cloned into MCS	This study
pRW224-*accB*-SC::*lacZ*	Promoter fragment cloned into MCS	This study

pRW225-*accB*-UTR-SC1::*lacZ*	Promoter fragment cloned into MCS	This study
pRW225-*accB*-UTR-SC2::*lacZ*	Promoter fragment cloned into MCS	This study
pRW225-*accB*-UTR-SC3::*lacZ*	Promoter fragment cloned into MCS	This study
pRW225-*accB*-UTR-SC4::*lacZ*	Promoter fragment cloned into MCS	This study
pRW225-*accB*-SC::*lacZ*	Promoter fragment cloned into MCS	This study
pRW224-*rplI*-p1.1::*lacZ*	Promoter fragment cloned into MCS	This study
pRW224-*rplI*-p1.2::*lacZ*	Promoter fragment cloned into MCS	This study
pRW225-*rplI*-p1.1::*lacZ*	Promoter fragment cloned into MCS	This study
pRW225-*rplI*-p1.2::*lacZ*	Promoter fragment cloned into MCS	This study
pBAD-*yhcB*-myc-His	*yhcB*-myc-His fusion downstream of an arabinose inducible promoter	This study
pBAD-*yhcB**	*yhcB* cloned downstream of an arabinose inducible promoter	This study

vortexed to mix for 15 s. The aqueous and organic phases were separated by centrifugation for 2 min at 13,000 x *g*. The DNA-containing aqueous layer was transferred to a new 1.5 ml microcentrifuge tube, 1/10th volume sodium acetate and 1/100th volume 1 M $MgCl_2$ was added to the DNA solution and mixed by vortex, followed by addition of 2 volumes of ice-cold 100% ethanol. The DNA was then concentrated by ethanol precipitation.

2.3.3 Ethanol purification of DNA

DNA samples extracted using phenol/chloroform extraction, or plasmid DNA too large to pass through Qiagen miniprep columns were purified and concentrated by ethanol precipitation. Samples, in ice-cold 100% ethanol, were incubated at -20°C for 1-2 h, or -80 °C for 20 min, then centrifuged for 15 min at 4°C at 13,000 x *g*. The supernatant was removed, and the remaining sample washed in 1 ml ice-cold 70% ethanol, then centrifuged again for 10 min at 4°C at 13,000 x *g*. The supernatant was removed and the remaining sample washed in 1 ml ice-cold 100% ethanol, then centrifuged for 10 min at 4°C at 13,000 x *g*. Finally, all supernatant was removed and residual ethanol allowed to evaporate. DNA was resuspended in 30 µl EB buffer (10 mM Tris-Cl, pH 8.5).

2.3.4 Qubit quantification of DNA

DNA was routinely quantified using Qubit™ dsDNA HS Assay kit (Invitrogen) following the product protocol at a 1:1 ratio of DNA to Dye using 1 µl of sample.

2.3.5 Tapestation quantification of DNA

DNA for sequencing was quantified using a Tapestation 2200 and the High Sensitivity D5000 kit (Agilent).

2.3.6 Polymerase chain reaction

The primers used in this study are listed in Table 2.4. Linear DNA fragments intended for cloning were amplified using Phusion High-Fidelity DNA Polymerase (New England Biolabs) and included the optional use of 5x GC buffer and 5% (v/v) DMSO. Linear fragments used for transformation in the Datsenko-Wanner method were amplified using Velocity DNA Polymerase (Bioline).

For colony PCR, a single colony was resuspended in 30 μl water, boiled for 10 min at 98°C, and debris pelleted at 11,000 x *g* for 1 min. 1 μl of supernatant was used as the PCR template with the MyTaq Red Mix (Bioline), half reaction volumes were used. Routine PCR thermal profiles are listed in Table 2.5.

2.3.7 PCR DNA purification

Standard PCR products amplified by Phusion and Velocity were purified using the Qiagen QIAquick PCR purification kit following kit instructions. Samples were routinely eluted in 35 μl nuclease-free water (Ambion), with an increased elution incubation time of 5 min and an additional elution step by re-incubating the eluted sample on the column to maximise DNA yield. For sequencing PCR clean-up and TraDIS sequencing library preparation, PCR reactions were purified using Solid Phase Reversible Immobilisation (SPRI) AMPure XP beads (Beckman Coulter) at a ratio of 0.9:1 beads to sample.

Table 2.4 Oligonucleotides used in this study

Name	Sequence (5′ to 3′)	Description
TnC_P1.F1	TCTTACGTGCCGATCAACGTCTCATTTTCGCC	Forward primer for amplification of the chloramphenicol transposon junction
TnC_.P1.R	GACTGGAGTTCAGACGTGTGCTCTTCCGATC	Reverse primer for amplification of the chloramphenicol transposon junction
TnC_P2_6.1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTACGGTCTCATTTTCGCCAAAGATGTGTA	CGTACG barcode
TnC_P2_6.3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTACGTAGTCTCATTTTCGCCAAAGATGTGTA	TACGTA barcode
TnC_P2_7.2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTAGCTGTCTCATTTTCGCCAAAGATGTGTA	GCTAGCT barcode
TnC_P2_7.4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGCTAGGTCTCATTTTCGCCAAAGATGTGTA	TAGCTAG barcode
TnC_P2_8.2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCATGCATGTCTCATTTTCGCCAAAGATGTGTA	GCATGCAT barcode
TnC_P2_8.3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCATGCATGGTCTCATTTTCGCCAAAGATGTGTA	CATGCATG barcode
TnC_P2_8.4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCATGCGTCTCATTTTCGCCAAAGATGTGTA	ATGCATGC barcode
TnC_P2_9.2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATCGATCGAGTCTCATTTTCGCCAAAGATGTGTA	ATCGATCGA barcode
TnC_P2_9.3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGATCGATGTCTCATTTTCGCCAAAGATGTGTA	TCGATCGAT barcode
TnC_P2_9.4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGATCGATCGTCTCATTTTCGCCAAAGATGTGTA	CGATCGATC barcode
TnK_P1.F	ACCTGCAGGCATGCAAGCTTCAGG	Forward primer for amplification of the kanamycin transposon junction
TnK_P2_6.1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTACGAGCTTCAGGGTTGAGATGTGTA	CGTACG barcode, kanamycin transposon
TnK_P2_6.3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTACGTAAGCTTCAGGGTTGAGATGTGTA	TACGTA barcode, kanamycin transposon
TnK_P2_7.2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTAGCTAGCTTCAGGGTTGAGATGTGTA	GCTAGCT barcode, kanamycin transposon
TnK_P2_7.4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGCTAGAGCTTCAGGGTTGAGATGTGTA	TAGCTAG barcode, kanamycin transposon
TnK_P2_8.2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCATGCATAGCTTCAGGGTTGAGATGTGTA	GCATGCAT barcode, kanamycin transposon
TnK_P2_8.3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCATGCATGAGCTTCAGGGTTGAGATGTGTA	CATGCATG barcode, kanamycin transposon
TnK_P2_8.4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCATGCAGCTTCAGGGTTGAGATGTGTA	ATGCATGC barcode, kanamycin transposon

TnK_P2_9.2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATCGATCGAAGCTTCAGGGTTGAGATGTGTA	ATCGATCGA barcode, kanamycin transposon
TnK_P2_9.3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGATCGATAGCTTCAGGGTTGAGATGTGTA	TCGATCGAT barcode, kanamycin transposon
TnK_P2_9.4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGATCGATCAGCTTCAGGGTTGAGATGTGTA	CGATCGATC barcode, kanamycin transposon
MfeI.CmTn.FO.F	GGGGCAATTGCTGTCTCTTATACACATCTACCGGGTCGAATTTGC	MfeI restriction site
MfeI.CmTn.RO.F	GGGGCAATTGCTGTCTCTTATACACATCTTTGGCGAAAATGAGACG	MfeI restriction site
HindIII.CmTn.FO1.R	GGGGAAGCTTCTGTCTCTTATACACATCTTTGGCGAAAATGAGACG	*Hin*dIII restriction site
HindIII.CmTn.RO1.R	GGGGAAGCTTCTGTCTCTTATACACATCTACCGGGTCGAATTTGC	*Hin*dIII restriction site
HindIII.CmTn.FO2.R	GGGGAAGCTTGCTGTCTCTTATACACATCTTTGGCGAAAATGAGACG	*Hin*dIII restriction site
HindIII.CmTn.FO3.R	GGGGAAGCTTGGCTGTCTCTTATACACATCTTTGGCGAAAATGAGACG	*Hin*dIII restriction site
HindIII.CmTn.RO2.R	GGGGAAGCTTGCTGTCTCTTATACACATCTACCGGGTCGAATTTGC	*Hin*dIII restriction site
HindIII.CmTn.RO3.R	GGGGAAGCTTGGCTGTCTCTTATACACATCTACCGGGTCGAATTTGC	*Hin*dIII restriction site
pRW_ins_check.F	AAGGACGAGAATTTCCCTGC	
pRW_ins_check.R	CCGTAATGGGATAGGTTACG	
ybbD_up.F	TCCATCAACAACAGATTGCTC	
ybbD_dn.R	AAGCAATTACTCTGCACAAGC	
ydaE_up.F	GGAAGACTTTAATCATGCGGT	
ydaE_dn.R	ATTTATGCCAGAAATGGCAGG	
yddK_up.F	CATCAGCTGCATAGTAGGG	
yddK_dn.R	CTATAACGATGAAGCCAACGG	
ydhL_up.F	TAGTAATACGCTGAACCATAATCG	
ydhL_dn.R	CAAGCAAAGTTGCTACTTTGG	
yedM_up.F	TTGAAGTCATTCGATGCGC	
yedM_dn.R	AACATCAGAGCGTCTCATTTG	
ymfE_up.F	AGCGAATATTAACTCCGTGCA	
ymfE_dn.R	AATATTGTCTGTCAGCGAAGC	
yncH_up.F	ATTTGCGTGGGATACCAGAAGA	
yncH_dn.R	TTAGCTCTGACTCATCAAGCT	
guaA_up.F	GTACTATCGACGAACTGCG	
guaA_dn.R	ATGTGGTGCAATCTGAAGC	
rpmF_up.F	ATTTGCCGTATTAGCCAGC	
rpmF_dn.R	CATGACATCTAACGCCAGG	
rpmI_up.F	TTACGCGAGCCATATGTGC	
rpmI_dn.R	CAGATGATCATGGTGCTCG	
ydcD_up.F	TTCTGCCAAATGTAGCGATG	
ydcD_dn.R	CCTTGCTCTCATGTCTTTCTG	
ykiB_up.F	TAAGTCTGTCATGGTCTGCC	
ykiB_dn.R	TTTCGCGCTAACGATTATGC	
ypjC_up.F	GTGATTCCATACAGACATTTGGC	
ypjC_dn.R	TCAGTAATTGTCTCAAATGAGCG	
ybgT_up.F	AACGTAACACCCACTCTCTG	
ybgT_dn.R	GCCCACATCAGCAATAAACC	
accB.promoter.EcoRI.F	GGGGGAATTCGTGAATTGTGCGGCTTGTTGC	*Eco*RI restriction site
accB.SC1.HindIII.R	GGGGAAGCTTAAACTGCGTCAGGGAGGACG	*Hin*dIII restriction site
accB.SC2.HindIII.R	GGGGAAGCTTTCCGCAGCGCAAAAACTGC	*Hin*dIII restriction site

accB.SC3.HindIII.R	GGGGAAGCTTCGTTGCGCCAATGTCACCT	*Hin*dIII restriction site
accB.SC4.HindIII.R	GGGGAAGCTTCTTGACGGACATTGTGCAGC	*Hin*dIII restriction site
accB.start.HindIII.R	GGGGAAGCTTACGAATATCCATGAGTGGGTTCCG	*Hin*dIII restriction site
rplI.promoter.EcoRI.F	GGGGGAATTCCGGTACCAGGAAGTTACGAGC	*Eco*RI restriction site
rplI.1.HindIII.R	GGGGAAGCTTACGGTCCATTAATACGACTTTGAG	*Hin*dIII restriction site
rplI.2.HindIII.R	GGGGAAGCTTGATCGCCATCAGTAATCGGTC	*Hin*dIII restriction site
yhcB_up.F	TAAGCGCCTTCAGGTATTGC	Forward primer for amplification of the *yhcB* locus
yhcB_dn.R	CCGACACTTAACGCTAATGC	Reverse primer for amplification of the *yhcB* locus
pBAD_ins_check.F	GCTATGCCATAGCATTTTTATCC	
pBAD_ins_check.R	CGTTCTGATTTAATCTGTATCAGG	

Table 2.5 PCR thermal profiles

MyTaq Red		
95°C	1 min	
98°C	15 s	35 cycles
55°C	15 s	
72°C	10 s/kb	
72°C	4 min	
4°C	hold	

Phusion		
98°C	30 s	
98°C	10 s	35 cycles
55°C	30 s	
72°C	30 s/kb	
72°C	10 min	
4°C	hold	

Velocity		
98°C	2 min	
98°C	30 s	35 cycles
55°C	30 s	
72°C	30 s/kb	
72°C	10 min	
4°C	hold	

2.3.8 Agarose gel electrophoresis

Isolated DNA were separated using 1% (w/v) agarose (Bioline) gels in 1x TAE buffer (50x TAE buffer: 2 M Tris base; 1 M acetic acid; 0.05 M EDTA in water) with 5 µl Midoori Green DNA dye (Nippon Genetics) per 100 ml agarose. DNA samples were mixed in a 1:1 ratio with DNA loading dye (0.025% (w/v) bromophenol blue; 0.025% (v/v) xylene cyanol F; 20% (v/v) glycerol; 10 mM Tris-HCl, pH 7.5; 1 mM EDTA) before loading into the wells. The DNA markers used were 100 bp or Hyperladder 1 kb DNA ladders (Bioline). Samples were separated by electrophoresis in 1x TAE buffer at 3-5 V/cm for 30-45 min. Gels were viewed and photographed under 300 nm ultraviolet light using a Gel Doc light box (Bio-Rad). However, DNA fragments to be purified by gel extraction were visualised using a blue light LED illuminator (Geneflow) to prevent DNA damage.

2.3.9 Extraction of DNA fragments from agarose gels

DNA samples were visualised as described in section 2.3.5. DNA bands were excised from the gel using a razor blade and transferred to a 1.5 ml microcentrifuge tube. DNA was extracted using the QIAquick Gel Extraction kit (Qiagen). DNA samples were eluted in 35 µl nuclease-free water (Ambion).

2.3.10 Restriction digestion of DNA

Purified PCR fragments and plasmids were digested independently with restriction enzymes (usually *Eco*RI and *Hin*dIII, FastDigest; New England Biolabs). DNA fragments were quantified using Qubit™ dsDNA HS Assay kit (Invitrogen). Digested

plasmids used for cloning were incubated with 2 µl (20 units) calf intestinal alkaline phosphatase (CIP; New England Biolabs) per 100 µl reaction at 37°C for 1 h.

2.3.11 DNA ligation

Vector and insert DNA for ligation were quantified using Qubit™ reagents as described in section 2.3.4. Reactions were set up on ice; vector and insert DNA were mixed at a molar ratio of 1:3, at a total amount of 50 ng of vector DNA. 2 µl of T4 DNA ligase buffer (10x) and 1 µl T4 DNA ligase (New England BioLabs) was added to the reaction and made up to a final volume of 20 µl with water. A reaction of cut vector only, without insert, was included as a ligation control. Reactions were incubated at 16°C for 16 h.

2.4 Bacterial transformation

2.4.1 Preparation of calcium competent cells

For 500 µl of competent cells, 50 ml of LB media was inoculated with a 1:100 dilution (500 µl) of overnight culture and incubated at 37°C with aeration until mid-exponential phase (OD_{600} = 0.4-0.6). The cells were stored on ice for 10 min then harvested by centrifugation for 10 min at 4,000 x *g* at 4°C. The supernatant was discarded and the cell pellet resuspended in 50 ml ice-cold 0.1 M $CaCl_2$. The cells were centrifuged and resuspended twice at 4,000 x *g*, 4°C for 10 min and resuspended in 25 ml ice-cold 0.1 M $CaCl_2$, then 5 ml ice-cold 0.1 M $CaCl_2$. The cells were centrifuged again at 4,000 x *g*, 4°C for 10 min and finally resuspended in 500 µl ice-cold freeze-thaw buffer (0.1 M $CaCl_2$; 15% (v/v) glycerol). Aliquots of 50 µl of competent cells were stored at -80°C.

2.4.2 Preparation of electrocompetent cells

To make competent cells, the initial volume of broth that cells were inoculated into was 100x the final volume required. For 500 μl of competent cells, 50 ml of LB media was inoculated with a 1:100 dilution (500 μl) of overnight culture and incubated at 37°C with aeration until mid-exponential phase (OD_{600} = 0.4-0.6). Cells were kept on ice for 30 min then centrifuged at 6,000 x *g* for 10 min at 4°C. The pellet was washed three times by resuspension and centrifugation with decreasing volumes of ice-cold distilled water (50 ml, 25 ml and 5 ml) and re-suspended in a final volume of 500 μl of ice-cold 15% (v/v) glycerol solution.

2.4.3 Transformation of competent cells

50 μl of cells were mixed with 50 ng of DNA in a 1.5 ml microcentrifuge tube and incubated on ice for 30 min. DNA was transformed into calcium competent cells by heat-shock: cells were incubated in a water bath at 42°C for 90 s then transferring to ice for 5 min. For electrocompetent cells, the cell and DNA solution was transferred to a 2 mm electroporation cuvette (Cell Projects) and pulsed at 2200 V using an Electroporator (Eppendorf). Cells were recovered in 1.5 ml microcentrifuge tubes with 1 ml SOC and incubated at 37°C with aeration, or 30°C if the vector is temperature sensitive, for 1 h. Cells were harvested by centrifugation at 4,000 x *g* and then resuspended in approximately 100 μl of the supernatant. Cells were spread on LB agar plates supplemented with the required antibiotic and incubated overnight at 37°C.

2.5 Chromosomal genetic engineering

2.5.1 P1 phage transduction

P1 transduction was used to transfer the kanamycin resistance cassette from the Keio collection of donor strains into *E. coli* strain BW25113 or into a strain constructed as part of this study (Fig. 2.1A). The protocol used was an abridged version of Thomason *et al.* (2007). Briefly, overnight cultures of both donor and recipient strain were prepared in LB with appropriate supplements if required. The overnight donor culture was diluted 1:100 in 5 ml LB plus 50 μl 20% (w/v) glucose and 25 μl 1M $CaCl_2$ in a 25 ml glass flask. Cultures were incubated 30-45 min at 30-37°C with shaking. 100 μl (10^9-10^{10} pfu/ml) P1 *vir* stock was added to the donor culture and incubated for a further 3 h at 37°C with shaking until cell lysis. 100 μl chloroform was added to the culture to complete cell lysis and kill the bacteria, and cultures were incubated for 10 min. Culture was centrifuged at 13,000 x *g* for 3 min, and the supernatant transferred to a glass bijou bottle for long term storage at 4°C. Recipient cells were prepared by centrifuging 1.5 ml overnight culture at 4,000 x *g* for 5 min to collect cells. The cell pellet was resuspended in 750 μl P1 salts (enough for 2 transductions; 10 mM $CaCl_2$, 5 mM $MgSO_4$). 4 glass bijou bottles were labelled per strain for P1 transduction: 1 each for 100 μl lysate, 10 μl lysate, 1 μl lysate and no lysate control. 100 μl recipient cells in P1 salts were added to the glass bijou bottles, followed by the lysate. Cells were incubated for 30 min at 37°C to allow phage to adsorb to the bacteria surface, then 1 ml LB and 200 μl sodium citrate is added to each bottle and the culture incubated for 1 h at 37°C with aeration. Cultures were centrifuged at 4,000 x *g* for 3 min, supernatant discarded and the cells re-suspended in the remaining supernatant. Cells were plated onto agar plates supplemented with 50 μg/ml kanamycin, and 5 mM

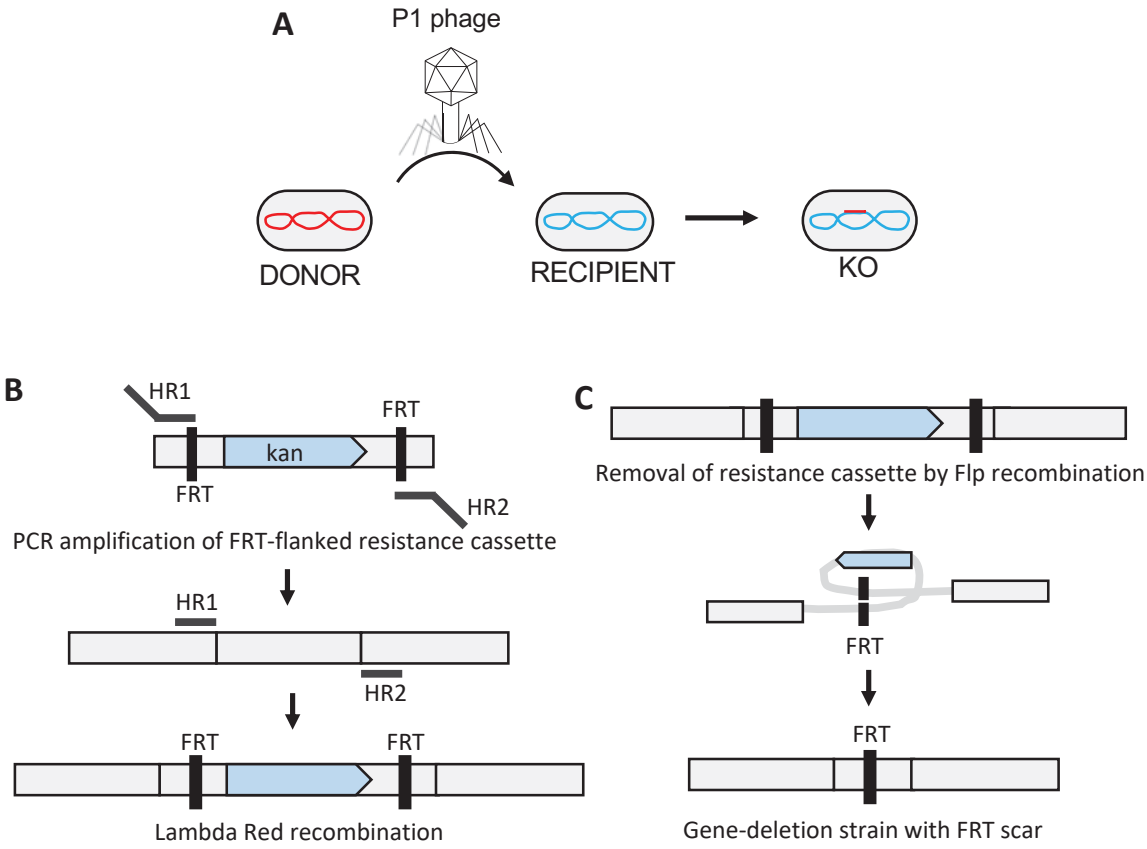

Figure 2.1 Chromosomal genetic engineering

(A) Schematic of P1 transduction of genetic material from donor to recipient strain. (B) Gene deletion via lambda-red recombination. (C) Removal of the resistance cassette via Flp recognition target (FRT) site recombination. Abbreviations: KO, knock out; HR, homology region; kan, kanamycin; PCR, polymerase chain reaction.

sodium citrate to prevent phage re-adsorption. 100 μl of lysate only control was also plated onto agar supplemented with kanamycin and sodium citrate as a negative control. Plates were incubated overnight at 37°C.

2.5.2 Datsenko and Wanner method for gene modification

The fragment for chromosomal insertion was amplified by PCR using pKD4 as a template, incorporating at least 50 bp of homology at either end of the template to the chromosomal insertion site (Fig. 2.1B). This fragment was then PCR purified and stored until required.

The target strain for gene modification was first transformed with the temperature sensitive plasmid pKD46, which codes for an arabinose inducible λ Red recombination system. The target strain carrying pKD46 was then grown at 30°C in LB supplemented with carbenicillin, and 0.2% (w/v) arabinose, and made electrocompetent following the steps in 2.4.2. 50 μl of competent cells were mixed with insert DNA in a 1.5 ml microcentrifuge tube and incubated on ice for 30 min. The Datsenko and Wanner method states transform 1 μg total DNA, but typically a range of volumes (1 μl, 5 μl and 10 μl) of purified PCR product were transformed (Datsenko & Wanner 2000). Cells were electroshocked with 2200 V, and transferred to a new 15 ml centrifuge tube with 1 ml SOC. Cells were recovered at 37°C with aeration for 2 h, then were then divided into equal aliquots of 500 μl. One aliquot was harvested by centrifugation at 4000 x *g* and then resuspended in approximately 100 μl of the supernatant. Cells were spread on LB agar plates supplemented with the required antibiotic and incubated overnight at 37°C. The second aliquot was left to recover at RT overnight for approximately 18 h before plating.

2.5.3 Kanamycin resistance cassette removal

Strains carrying a kanamycin resistance cassette for removal were first transformed with pCP20 and incubated at 30°C (as plasmid replication is temperature sensitive). The pCP20 plasmid codes for the yeast recombinase Flp which recombines specific target sequences, 'Flp recombination target' (FRT) sites (Cherepanov & Wackernagel 1995). The kanamycin resistance cassette contains two FRT sites either side of the resistance gene, oriented in the same direction. The Flp recombinase expressed from the pCP20 plasmid mediates recombination between the two FRT sites (Fig. 2.1C) removing the kanamycin resistance cassette and leaving a 102 bp scar. To lose the plasmid, colonies were inoculated in 5 ml LB and grown for ~6 h at 37°C, then streaked onto agar plates and incubated overnight at 37°C. Individual colonies were selected and patch plated onto LB agar plates, LB agar supplemented with kanamycin (to confirm deletion) and LB agar supplemented with carbenicillin (to confirm plasmid loss). Colonies that grew on only the LB agar plate were then checked by colony PCR to confirm deletion.

2.6 Promoter reporter assays

2.6.1 β-galactosidase assay

β-galactosidase assays were used to measure the activity of promoter constructs cloned into pRW224 or pRW225 plasmids upstream of the *lacZ* gene. Strains carrying the relevant plasmid with a promoter::*lacZ* fusion were grown overnight at 37°C with aeration in LB and supplemented with 35 μg/ml tetracycline. Density of the overnight culture was determined by measuring OD_{650} and used to sub-culture into 5 ml of LB supplemented with 35 μg/ml tetracycline to a final OD_{650} of ~0.025. Cultures were

incubated at 37°C with aeration until mid-exponential phase (OD_{650} = 0.3-0.5). At this point, the end OD_{650} was recorded and 100 µl each of toluene and 1% sodium deoxycholate was added to the culture to lyse cells. After vortexing, cultures were returned to the incubator for 30 min at 37°C. Lysates were stored in the fridge if needed until all cultures to assay were ready.

The β-galactosidase activity of each culture was measured by adding 100 µl lysate to 2.5 ml Z buffer (10 mM KCl; 1 mM $MgSO_4 \cdot 7H_20$; 60 mM Na_2HPO_4; 30 mM $NaH_2PO_4 \cdot 2H_2O$) supplemented with 270 µl 2-Hydroxyethylmercaptan (β-mercaptoethanol, Sigma Aldrich) and 80 mg ortho-Nitrophenyl-β-galactoside (ONPG, Thermo Scientific) per 100 ml Z buffer. The reaction was incubated at 37°C and the time point of lysate addition was recorded. The reaction was stopped when a pale-yellow colour developed, or if 120 min had passed with no change, by adding 1 ml 1 M sodium carbonate. The end time point of the reaction was also recorded and absorbance of the reaction at OD_{420} was measured using a "no lysate" control as a blank to subtract background absorbance. β-galactosidase activity was measured in Miller units using the following calculation:

$$\beta\text{–galactosidase activity} = \frac{1000 \text{ x } 2.5 \text{ x } 3.6 \text{ x } OD420}{4.5 \text{ x } OD650 \text{ x } t \text{ x } v} \; nmol/min/mg \; bacterial \; mass$$

Each strain was assayed with 3 biological replicates. The mean and standard deviation was calculated for each strain.

2.7 Preparation of cellular fractions

2.7.1 Whole cell lysate protein extraction

1 ml of overnight culture diluted to an OD_{650} of 1.0 was transferred to a 1.5 ml microcentrifuge tube and centrifuged at 13,000 x *g* for 1 min. The supernatant was discarded and cell pellet resuspended in 100 µl Laemmli sample buffer (Sigma Aldrich). The sample was then incubated at 98°C for 10 min to lyse cells and centrifuged at 13,000 x *g* for 1 min to collect debris.

2.7.2 Total phospholipid extraction

Total lipid from whole cell samples was extracted using an amended version of the Bligh-Dyer method (Bligh & Dyer 1959). 10 ml of overnight culture was centrifuged at 3,220 x *g* for 10 min at 4°C. The supernatant was discarded and pellet resuspended in 1 ml ddH_2O and transferred to a glass tube. 1.25 ml chloroform and 2.5 ml methanol were added to the sample using glass pipettes. The sample was vortexed for 20 s to create a single phase solution which was then incubated at 50°C for 30 min. A further 1.25 ml chloroform and 1.25 ml water was added to the sample to create a 2-phase solution and incubated again at 50°C for 30 min. The sample was centrifuged 400 x *g* for 10 min at RT and the lower organic phase containing phospholipids was transferred to a new glass tube using a glass Pasteur pipette. Chloroform was evaporated by placing the glass tubes in a heat block at 50°C under a stream of nitrogen. Dried samples were either stored at -20°C or washed once more with 7.25 ml chloroform:methanol:water in a 1:1:1 (v/v) ratio and incubated 30 min at 50°C before centrifuging and drying.

2.8 Protein analysis

2.8.1 Protein induction

Gene expression from a pBAD vector was induced typically by adding arabinose to growth media at a final concentration of 0.4%. A concentration of 0.2% (w/v) glucose was used to repress expression for the P_{BAD} promoter.

2.8.2 SDS-PAGE analysis

Typical protein samples were analysed on either a 10% or 15% resolving gel with a 4% stacking gel made in a 1 mm gel cast (BioRad). For 15 ml of resolving gel, a 10% gel comprised: 6.15 ml water; 3.75 ml 1 M Tris-HCl, pH 8.8; 0.075 ml 20% (w/v) SDS; 5 ml Acrylamide/Bis-acrylamide (30%/0.8% w/v, Protogel); 0.075 ml 10% (w/v) ammonium persulfate (APS); and 0.01 ml TEMED. A 15% gel comprised: 3.6 ml water; 3.75 ml 1 M Tris-HCl, pH 8.8; 0.075 ml 20% (w/v) SDS; 7.5 ml Acrylamide/Bis-acrylamide (30%/0.8% w/v, Protogel); 0.075 ml 10% (w/v) APS; and 0.01 ml TEMED, adding the APS and TEMED last. Per 5 ml 4% stacking gel there were 3.075 ml water, 1.25 ml 0.5 M Tris-HCl (pH 8.8), 0.025 ml 20% (w/v) SDS, 0.67 ml Acrylamide/Bis-acrylamide (30%/0.8% w/v, Protogel), 0.025 ml 10% (w/v) APS and 0.005 ml TEMED.

Samples were run in 10 or 15 well gels, loading a maximum of 20 or 10 µl sample respectively. Gels were run initially at 100 V until the sample left the stacking gel, and then at 120 V until the loading dye reached the bottom of the gel. Gels were stained using a mix of 20% (v/v) isopropanol, 0.4 M citric acid and 2% (w/v) Coomassie R for 30 min and destained using water.

2.8.3 Western blotting

Samples for western blotting were separated by SDS-PAGE as outlined above. Protein was transferred from the gel to nitrocellulose membrane using an iBlot2 (Invitrogen). Following protein transfer, the membrane was first blocked using a milk buffer (per 500 ml: 25 g skim milk powder and 0.5 g sodium azide in Tris-buffered saline [TBS: 8 g NaCl; 1.2 g Tris-base, pH 7.4; per L]) for 1 h at RT to prevent background antibody binding. The membrane was then incubated in primary antibody (diluted in milk solution) overnight at 4°C. The membrane was washed three times in TBST for 5 min (per L: 1 ml Tween-20, 1.2 g Tris-base, 8 g NaCl; pH 8.4). The wash was then poured off, and the membrane incubated in secondary antibody conjugate to alkaline phosphatase, and incubated for 1 h at RT. The western blot was developed using substrate NBP-BCIP (nitroblue tetrazolium chloride–5-bromo-4-chloro-3′-indolylphosphate; Sigma Aldrich).

2.9 Lipidomics techniques

2.9.1 Thin layer chromatography

10 cm of TLC silica gel membrane (Merck) was cut from a template, ensuring not to gouge the membrane. 5-10 µl of sample was spotted onto the origin using a 5 µl glass capillary tube (Sigma Aldrich). Once dry, the membrane was transferred to an equilibrated solvent system of 65:25:10 chloroform:methanol:acetatic acid. The samples were separated over 30 to 45 min until the solvent front had migrated approximately 9 cm from the origin, then the membrane was removed from the solvent tank and was air dried at RT. Samples were stained with phosphomolybdic

acid (PMA) and heated with a heat gun to activate the PMA until lipid species were visible. Samples were imaged using a GS-800 densitometer (Bio-Rad).

2.10 Phenotypic screening

2.10.1 Microdilution spot plate

LB agar plates were set as described in section 2.1.1 and supplemented with the desired antibiotic or antimicrobial when required. Overnight cultures of strains were diluted to 1 ml at an OD_{600} of 1.00. 200 µl of culture was then transferred to row A of a 96-well microtiter plate and 10-fold serially diluted by transferring 20 µl into 180 µl LB in row B, and repeated as far as row H. 2 µl of culture from each well was inoculated onto the plate using a multi-channel pipette and allowed to dry before plates were incubated overnight at 37°C. Images were recorded using a G:BOX light box (Syngene).

2.10.2 Gradient plate

Gradient plates were poured in square plates by first setting the plate on a slant (using the plate lid as a support underneath the edge of the plate) and pouring the first layer of 40 ml medium. The bottom layer was typically the antibiotic-supplemented layer. Once fully set, the plate was then placed flat on the bench and a fresh layer of 40 ml LB agar poured on top. The plate was left overnight to allow compounds to equilibrate before inoculating. To inoculate a gradient plate, cells were diluted to an OD_{600} of 0.10 and 2 µl were inoculated across the width of the plate (12 spots total) and allowed to dry before incubating overnight at 37°C.

2.11 Microscopy

Strains were grown to mid log phase OD_{600} 0.5 from a 1:100 dilution of overnight culture in 5 ml LB, or taken directly from overnight samples and diluted to an OD_{600} of 0.10. 5 µl cells were spread on a 1 mm glass slide pre-treated with 5 µl poly-L-lysine (Sigma Aldrich). A Nikon 90i eclipse microscope was used to capture differential interference contrast (DIC) images of cells, using the 100x or 40x objective lens with a Nikon immersion oil

2.12 TraDIS method

The TraDIS method for preparing samples for sequencing used in this thesis was devised and optimised by Ashley Robinson, based on the techniques of other published works (Langridge *et al.* 2009; Phan *et al.* 2013; Christen *et al.* 2014). The *E. coli* chloramphenicol transposon library with a chloramphenicol resistance cassette used in this study was constructed by collaborator Keith Turner. This transposon library has a mini-Tn*5* transposon carrying a chloramphenicol resistance cassette, amplified from pACYC184, in strain BW25113 (Chang & Cohen 1978).

2.12.1 Transposon library construction

The transposon library in BW25113 Δ*yhcB* strain was constructed using the following method. Deletion of the *yhcB* gene was confirmed by PCR. The strain was also grown on LB supplemented with kanamycin, and LB supplemented with carbenicillin to confirm excision of the kanamycin resistance cassette and loss of the pCP20 plasmid before library construction. 10 ml of 2x TY broth was inoculated with a single colony of the BW25113 Δ*yhcB* strain, and grown overnight at 37°C with

aeration. The 10 ml overnight culture was added to 800 ml 2x TY broth in a 2 L flask, and grown at 37°C with aeration until OD_{600} 0.9. At the desired OD, 600 ml of culture was distributed between 12x 50 ml centrifuge tubes and incubated on ice for 30 min. Cells were then centrifuged at 5,000 x *g* for 10 min at 4°C. The supernatant was discarded, and 5 ml ice-cold 10% (v/v) glycerol was gently added to the centrifuge tube. Cells were resuspended gently using a 1 ml pipette, and samples diluted to 50 ml 10% (v/v) glycerol. Cells were centrifuged again at 5,000 x *g* for 10 min at 4°C, and resuspended using the above method but to a final volume of 25 ml ice-cold 10% (v/v) glycerol. Two aliquots of 25 ml culture were combined, and the process of centrifugation and resuspension in 50 ml of 10% (v/v) glycerol, followed by 25 ml of 10% (v/v) glycerol, was repeated until all 12 samples were combined and evenly distributed between 2x 50 ml centrifuge tubes. The final two centrifuge tubes were centrifuged at 5,000 x *g* for 10 min at 4°C. One pellet was first resuspended gently in 1 ml of ice-cold 10% (v/v) glycerol. This sample was then transferred to the second pellet and used to resuspend the final pellet in a final volume of ~2 ml. Aliquots of 200 μl cells were distributed between 1.5 ml microcentrifuge tubes. 0.2 μl EZ-Tn5™ transposome (Epibio) was mixed with each aliquot of cells and incubated on ice for 30 min. Samples were transferred to pre-chilled 2 mm gap electroporation cuvettes (Cell Projects Ltd.). Cells were pulsed at 2200 V and 2 ml of pre-warmed SOC medium was immediately added to the sample for recovery. The sample was transferred to a 15 ml falcon tube to allow for maximum aeration, and was incubated at 37°C for 2 h. 5 ml of LB broth was added to each 15 ml falcon tube resulting in ~10 tubes each with ~6 ml of sample. Using a 10 ml stripette, ~4-5 drops of the sample (equivalent to ~200 μl) were spread per LB agar plate supplemented with 50 μg/ml kanamycin.

Sufficient cells were inoculated per plate to form non-touching single colonies, no more than 2,000 per plate. Plates were incubated overnight at 37°C.

Following incubation, 500 µl LB broth was added to each plate and using a 'hockey-stick' spreader, colonies were scraped off the surface of the agar plate and resuspended in the broth medium. Cells were transferred to a 50 ml falcon. When 45 ml of culture was acquired, 5 ml of sterile 100% glycerol was added to the falcon tube. Cells were mixed thoroughly before storing at -80°C.

If multiple libraries were made in the same parent strain, these were thawed on ice (if necessary), mixed thoroughly and aliquots of each falcon collected to form a pooled library, which was then thoroughly mixed again and 1 ml aliquots extracted into cryotubes for easy access.

2.12.2 Determination of a relative MIC concentration in broth for TraDIS experiments

The minimum inhibitory concentration (MIC) of polymyxin B used for TraDIS experiments was determined using an amended version of the Andrews broth microdilution protocol (Andrews 2001). The conditions were changed to reflect the conditions of a TraDIS growth protocol. The starting inoculum was increased to an initial OD_{600} of 0.05 (in line with the TraDIS growth starting inoculum), and the medium used was LB.

2.12.3 TraDIS library screening and DNA isolation

DNA was either extracted directly from the transposon library or harvested from a grown culture. For LB growth studies, 50 ml LB was inoculated with 10 µl of library

(OD_{600} ~0.05) and grown at 37°C with aeration, 5 ml of culture was harvested by centrifugation at 4,000 x *g* for 10 min at 4°C, at OD_{600} = 1.0, 2.0, 3.0 and 4.0. For polymyxin B screens, 50 ml LB was inoculated with 10 µl of library (OD_{600} ~0.05) and supplemented with the desired concentration of polymyxin B (Fig. 2.2). The culture was grown until an OD_{600} of 1.00 then was harvested by centrifugation at 4,000 x *g* for 10 min at 4°C. Genomic DNA was extracted from the cell pellets using the STRATEC RTP Bacteria DNA Mini kit as outlined in section 2.3.1.

2.12.4 DNA fragmentation

To process genomic DNA for sequencing a series of steps are followed, outlined in figure 2.3. Genomic DNA was quantified using the Qubit method as outlined in section 2.3.4. 1 µg of DNA was diluted in a final volume of 500 µl Nuclease-Free Water (Ambion) and incubated on ice for 10 min. A bioruptor (Diagenode) was used to shear DNA samples mechanically by pulses of ultrasound waves through an ice-cold water bath. The shearing profile used was 30 s ON, 90 s OFF at low intensity to fragment DNA to an average length of ~300 bp. Following shearing, samples were condensed using a vacuum concentrator (Eppendorf, concentrator 5301) to a final volume of 55.5 µl.

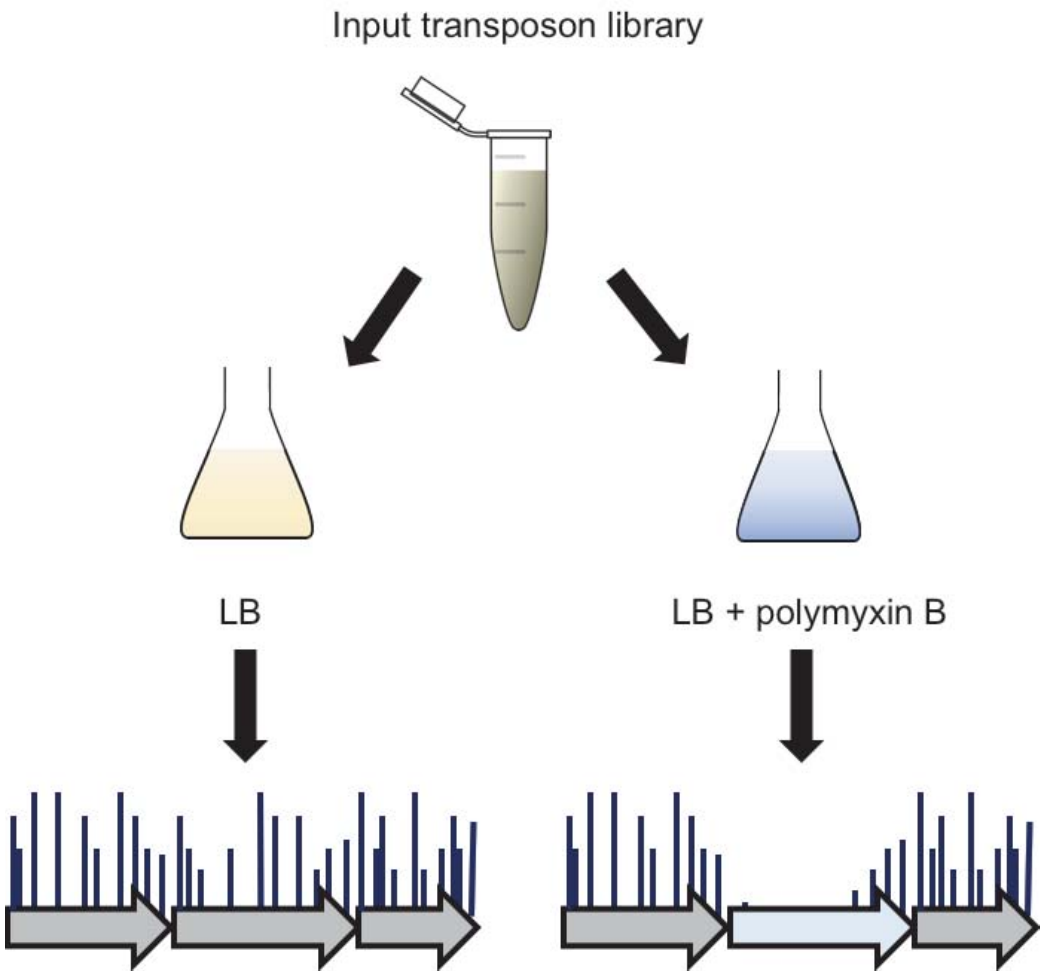

Figure 2.2 Overview of the protocol for the identification of conditionally essential genes

Growth of the transposon library with and without polymyxin B to identify conditionally essential genes required for survival in the presence of the antibiotic. Genes required for growth in the presence of polymyxin B (blue), if disrupted, will result in a mutant that is outcompeted during growth and therefore underrepresented within the population following sequencing.

2.12.5 Sequencing library preparation

NEBNext Ultra I kit (New England Biolabs) was used to repair the ends of fragmented DNA and ligate an adaptor to the newly prepared DNA fragment ends, following the kit instructions. After the excision of uracil in the adaptor hairpin loop, the sample was purified by doing a two-sided size selection step using AMPure XP SPRI beads (Beckman Coulter) to select for fragments of 250 bp following the kit instructions, except the final elution step where the sample was eluted in 16 μl of nuclease free water. Next was a custom step that was introduced to enrich for transposon junction fragments. The enrichment PCR step amplified DNA between a custom forward primer that annealed to the transposon end and a reverse primer that annealed to the ligated adaptor, to amplify the transposon-gDNA junction (Fig. 2.3). The PCR reaction contained 25 μl 2x KAPA HiFi DNA Polymerase (Kapa Biosystems), 2.5 μl TnC/K_P1.F1 (10 μM), 2.5 μl TnC_P1.R (10 μM), 15 μl sample and 5 μl nuclease-free water and was amplified using the following cycle conditions: 98°C, 48 s; (98°C, 15 s; 65°C, 30 s; 72°C, 30 s) x10; 72°C, 1 min; 4°C hold. The PCR product was purified as outlined in section 2.3.6 using SPRI beads at a ratio of 0.9:1 beads to sample. The next PCR step prepared the sample for sequencing through the addition of flow-cell adaptors to allow the sample to bind to the MiSeq flow cell and provide a priming site for the sequencing primers, in addition to multi-plexing the samples and staggering the transposon sequence start to introduce variation in the early cycles of sequencing (Fig. 2.4). The PCR reaction contained 25 μl 2X HiFi DNA Polymerase (KAPA), 2.5 μl custom forward primer (TnC/K_.X.Y; 10 μM), 10 μM NEBNext Index Reverse primer, 15 μl sample made up to a total reaction volume of 50 μl with nuclease-free water. Finally, the PCR product was purified as outlined in section 2.3.6 using SPRI beads at

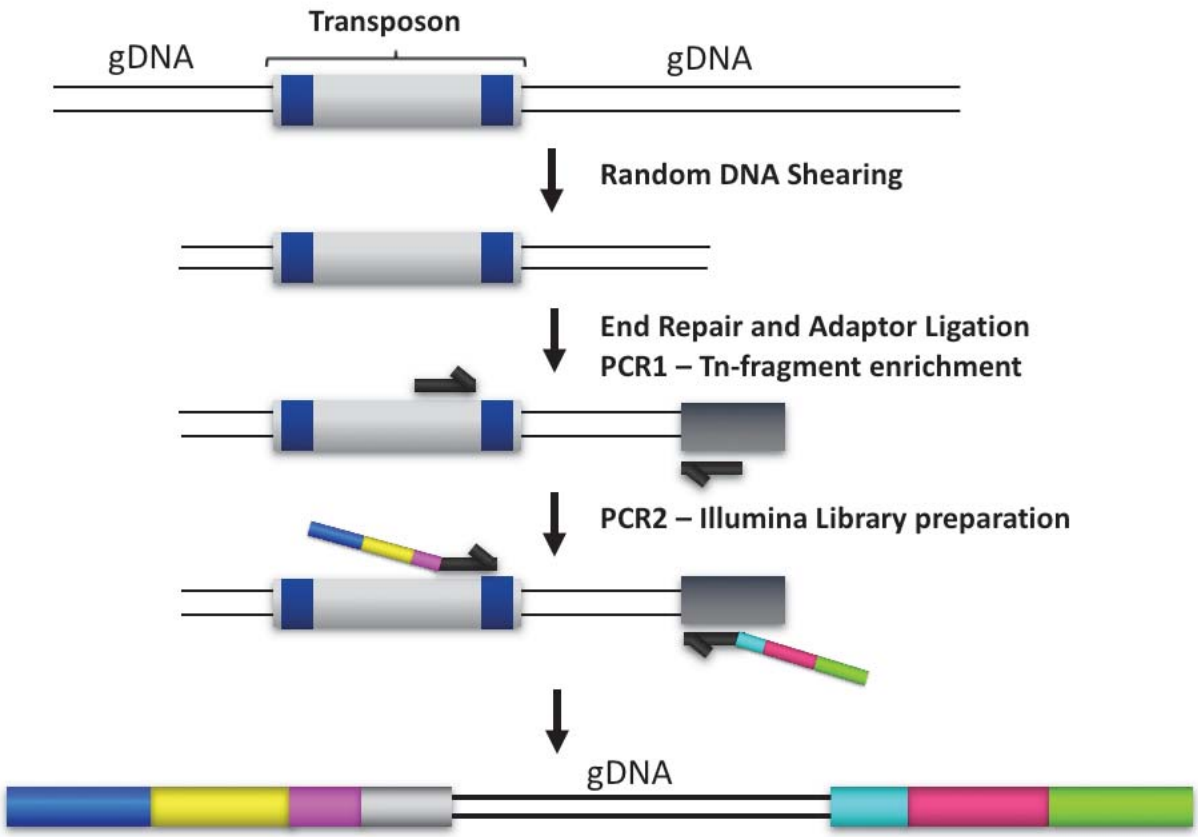

Figure 2.3 Preparing genomic DNA for sequencing

Genomic DNA is prepared for sequencing first by fragmentation; the fragment ends are repaired and adaptor ligated. The transposon-gDNA junction is amplified by PCR with primers specific for the transposon and ligated adapter. The fragments are then prepared for sequencing by a second PCR step that introduces sequencing specific adapters and barcodes.

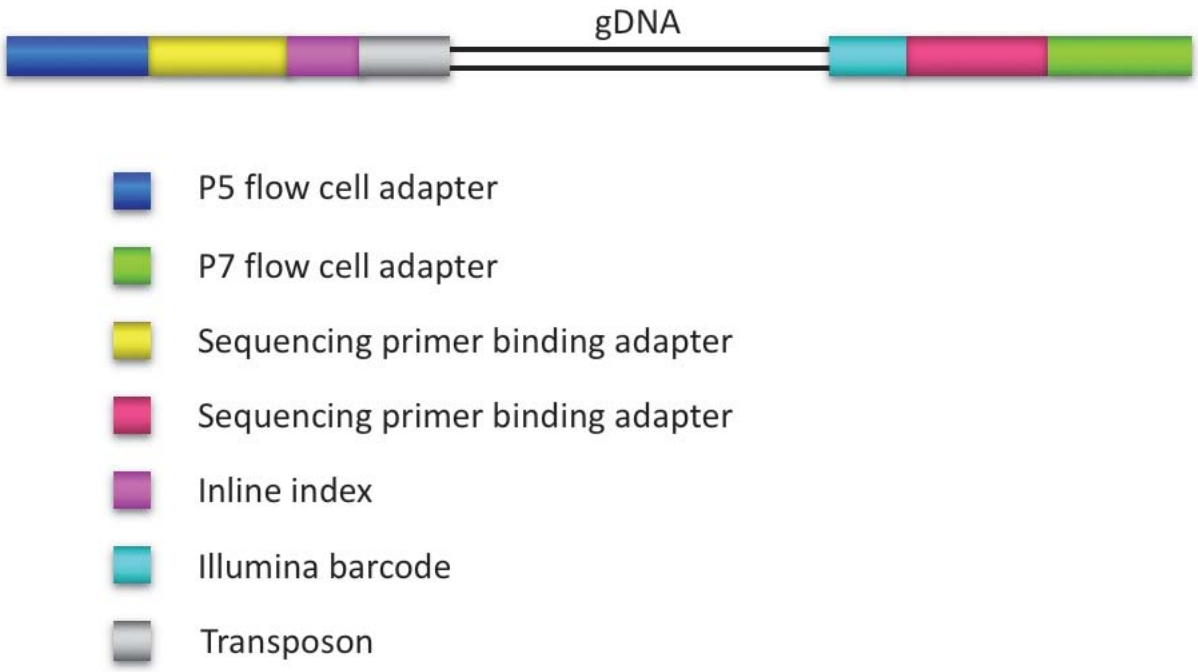

Figure 2.4 Schematic of a DNA fragment ready for sequencing

A fragment for sequencing contains at the extreme ends of the fragment a P5 and P7 adapter required for binding to an Illumina flow cell (blue and green, respectively). Next, are Illumina adapters which provides the sequencing primer binding sites (yellow and red). Inside from the adapters are two barcodes at either end of the fragment: the inline index barcode which staggers the introduction of the transposon start (pink), and the Illumina barcode at the opposite end to identify the sample (cyan). The innermost portion of sample is the transposon-gDNA junction.

a ratio of 0.9:1 beads to sample and eluted in 33 μl nuclease-free water. 32 μl sample was transferred to a new 1.5 ml microcentrifuge tube and stored at -20°C unless sequencing within a week. The final product ready for sequencing is outlined in figure 2.4.

2.12.6 Quantification of sequencing libraries using qPCR

Genomic libraries ready for Illumina sequencing were quantified using the KAPA Library Quant Kit (Illumina) Universal qPCR Mix (Kapa Biosystems), following the kit instructions at half volume (10 μl) reactions. As standard, 3 independent replicates of the library were quantified at 2 different dilutions: 1/50,000 and 1/500,000, and a no template control was used as a blank control. Samples were quantified using a Mx3005P qPCR system (Agilent Technologies) and the thermal profile used was: 95°C for 5 min, then 35 cycles of 95°C for 30 s followed by 60°C for 30 s, including a dissociation melt curve from 65°C to 95°C to assess sample quality.

2.12.7 MiSeq sequencing of the transposon junction

The concentration of the library calculated by qPCR was used to dilute all samples to 8 nM stock concentrations. 1.5 μl of each 8 nM sample was transferred to a nuclease-free 1.5 ml microcentrifuge tube to combine and pool the samples. The library was then denatured and diluted following Illumina protocols, and spiked with 5% denatured PhiX library (20 pM). The sample was loaded on the MiSeq (Illumina) to aim for an optimal cluster density of 1,000 clusters per mm^2 using 150 cycle v3 cartridges.

2.12.8 Sequence data analysis

The sequencing output .fastq files were processed using a customised version of the scripts originally written by Ashley Robinson and detailed in his thesis (Robinson, 2016). The data processing steps are outlined in Fig. 2.5. Briefly, sequencing reads were first automatically filtered by the Illumina barcode, then further separated by the inline-index barcode used per replicate, using the Fastx barcode splitter and trimmer tools (Pearson *et al.* 1997). Surviving reads were checked in two steps for an accurate transposon sequence: the data were checked first for 25 bp of transposon sequence introduced by the TnC_P2_X.Y primer, or (depending on the transposon used) 23 bp of transposon sequence introduced by the TnK_P2_X.Y primer (X = 6-9, Y = 1-4) allowing for up to 3 nucleotide base mismatches. If there was a match for the first transposon tag, the sequence was checked for the last 10 bp of transposon allowing up to 1 base mismatch. Correctly identified transposon sequence was trimmed using Trimmomatic (Bolger *et al.* 2014) and parsed to the next step. Poor quality data and reads <20 bp were removed. Reads were then mapped to the *E. coli* BW25113 reference genome (CP009273) using the bwa aligner mem algorithm (0.7.8-r455; Li & Durbin 2009). Reads were sorted and indexed using samtools (0.1.19-44428cd) and converted to a .bed file (using the bedtools suite), which was then intersected against the annotated protein coding sequences in the BW25113 general feature format (.gff) file downloaded from NCBI (Li & Durbin 2009; Quinlan & Hall 2010). The location and frequency of transposon insertion events was quantified using custom python scripts. Data were inspected manually using the Artemis genome browser (Rutherford *et al.* 2000).

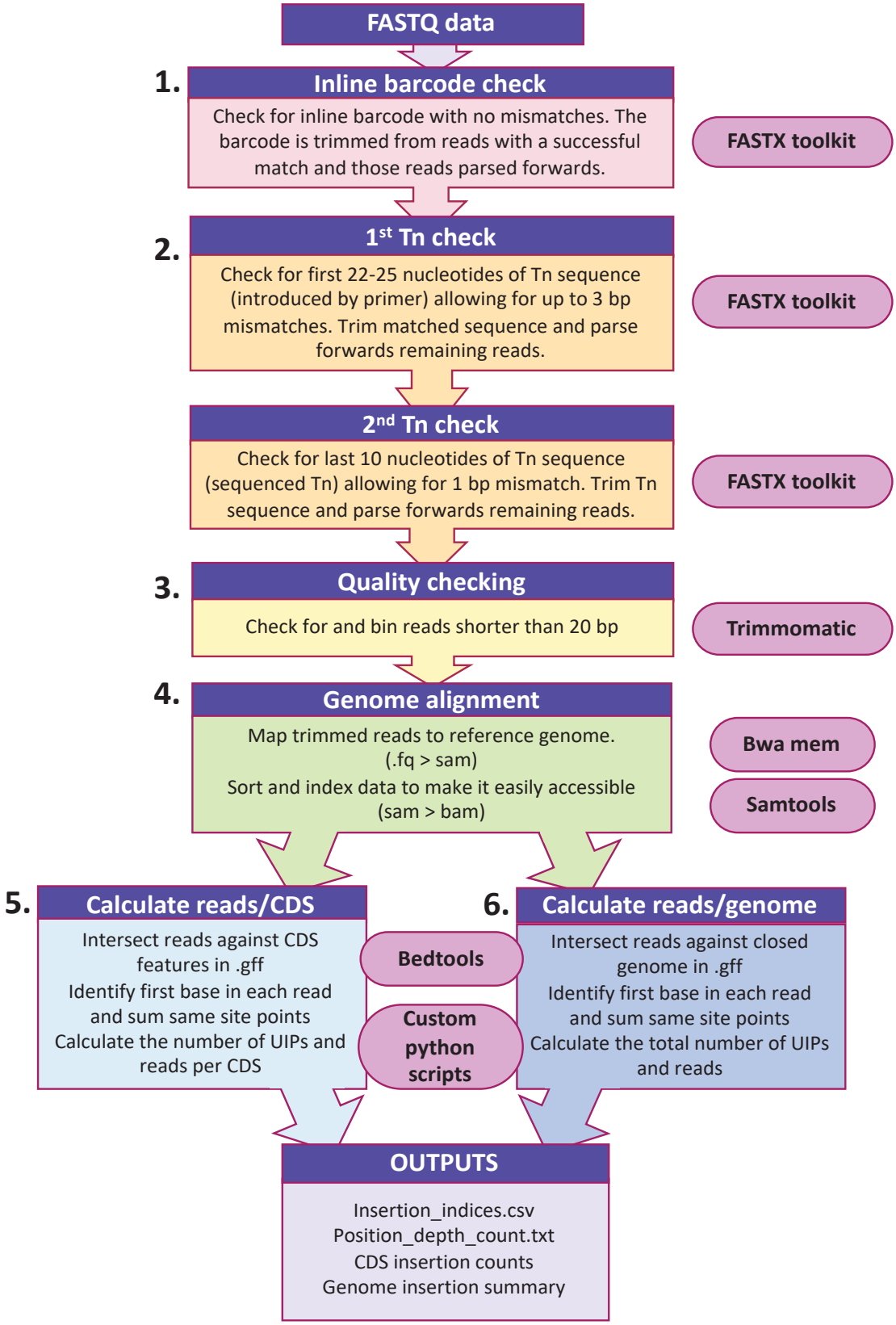

Figure 2.5 Schematic of the TraDIS data processing pipeline

Data that survives each process is passed to the next stage. Tools used in each step are shown in pink boxes to the side of the described process

2.12.9 Essential gene prediction

Essential gene prediction for the chloramphenicol transposon library (TL) and LB datasets were done by Sara Jabbari using an amended version of the scripts kindly given by the authors of Langridge *et al.* (2009). Essential gene prediction for subsequent datasets was done by the author. The Freedman-Diaconis method was used to generate a histogram with data-informed bin widths. Using the R MASS library (http://www.r-project.org), an exponential distribution was fitted to the left, "essential" mode; a gamma distribution was fitted to the right, "non-essential" mode. The probability of a gene belonging to each mode was calculated and the ratio of these values was used to calculate a log-likelihood score. Using a 12-fold likelihood threshold, based on the log-likelihood scores, genes were assigned as "essential" if they were 12 times more likely to be in the left mode than the right mode, and "non-essential" if they were 12-times more likely to be in the right mode than the left mode. Genes with log-likelihood scores between the upper and lower $\log_2(12)$ threshold values of 3.6 and -3.6 respectively were classed as "unclear".

2.12.10 Essential gene lists

The Keio essential gene list consists of the original essential genes minus three ORFs, JW5190, JW5193 and JW5379, as they are not annotated within MG1655 and are thought to be spurious, resulting in a final list of 300 genes (Baba *et al.* 2006; Zhou & Rudd 2013). The PEC (profiling of the *E. coli* chromosome) dataset consists of the 300 genes listed as essential for W3110 (Yamazaki *et al.* 2008). The lists of essential genes were compared using BioVenn (Hulsen *et al.* 2008).

2.12.11 Statistical analysis of insertion density

Statistical analysis of the likelihood of an insertion-free region was done by Iain Johnston. A geometric distribution was used to model the probability of finding an insertion free-region (IFR). Following this distribution, there are three probabilities to consider: (i) a single region of length '*L*' has no insertions; (ii) a gene of length g has one or more IFRs of length *L* (ρ_{gene}); (iii) a genome of length G has one or more IFRs of length *L* (ρ_{genome}). The probability of seeing k-1 "failures" (no insertion) followed by a "success" (an insertion) in a string of independent trials is $P(k) = \rho(1-\rho)^{(k)}$, where ρ is the probability of an insertion. Therefore, for an IFR of length *L*, the *P* value is $P = \sum_{k=L}^{\infty} P(k)$. Calculation of the true ρ_{gene} and ρ_{genome} values were approximated computationally for simplicity. Random transposon insertion was simulated N times in a genome of size G to determine the statistics of the resulting IFRs.

2.13 Sequencing

2.13.1 Whole genome sequencing

Whole genome sequencing was done by MicrobesNG, University of Birmingham, UK. Cultures for sequencing were streaked to single colonies on an appropriate agar plate, a single colony was selected and resuspended in 100 μl PBS. The PBS culture was spread to cover half of an appropriate agar plate to create a lawn, with a small sample streaked to single colonies to check for contamination. Following growth, the lawn was collected and transferred to a bead tube provided by MicrobesNG and sent to the sequencing facility.

2.13.2 Sanger sequencing

Plasmid and PCR products were sequenced by the Functional Genomics and Proteomics Laboratory, University of Birmingham, UK. Plasmid templates were extracted using the Qiagen QIAprep Spin Miniprep kit as outlined in section 2.3.1. 100 ng of plasmid DNA was mixed with 1 çM of sequencing primer in a final volume of 10 µl. PCR product was purified using the Qiagen QIAquick PCR purification kit as outlined in section 2.3.5. 1-100 ng (fragment size-dependant) of purified product was mixed with 1 µM of sequencing primer in a final volume of 10 µl.

2.14 Bioinformatics

2.14.1 Genome alignment

Illumina sequencing reads of sequence data were aligned to reference genome BW25113 available from the NCBI database (CP009273.1) using bwa mem and then converted to bam files and sorted and indexed using SAMtools (Li *et al*. 2009; Langmead & Salzberg 2012).

2.14.2 Genome assembly

Raw Illumina fastq files were processed and aligned using SPAdes (3.10.1; Bankevich *et al*. 2012). Assembly data were checked using Quast (Gurevich *et al*. 2013).

2.14.3 Identification of chromosomal variations

The programs Snippy and VarScan were used to identify SNPs and indels. The program BreSeq was used to identify large chromosomal rearrangement events (Koboldt *et al*. 2012; Deatherage & Barrick 2014; Seeman 2015).

2.14.4 Data handling

General data management, file conversion and data sub-setting was done using R studio (http://www.r-project.org), and using the dplyr and tidyr libraries, and custom scripts, or using python. The ggplot library was used for creating figures in R, and the matplotlib library was used for creating figures in python.

Chapter 3

The Essential Genome of E. coli K-12 Identified using TraDIS

Declaration

The focus of this chapter is analysis of data generated by a previous student, Ashley Robinson, and a continuation of work presented in his thesis. The initial TraDIS experiments (growth and sequencing of the input and output mutant pools) with the library (constructed by Keith Turner) were completed by Ash Robinson. The TraDIS protocol and processing pipeline were also devised by Ash Robinson. The data collected by Ash Robinson, were mapped to the BW25113 genome by me. The subsequent analysis presented in this chapter was completed by me, with the exception of statistical analysis, as clearly stated in the text. Specifically, I processed the raw fastq data (using Ash's scripts) mapping the data to the BW25113 genome. Sara Jabbari identified the essential genes, which I then compared to the Keio and PEC lists of essential genes. I also wrote the scripts to convert the data into figures. I prepared the results of this analysis for publication and I wrote the text of the publication, which was edited by senior co-authors (Goodall *et al.* 2018). The sections of the text that I did not write (specialist statistical analysis) have not been included in this chapter. Much of the text in this chapter is that in the final publication, however additional data are presented together with data from further TraDIS experiments completed by me.

E. C. A. Goodall
Co-Author

A. Robinson
Co-Author

I. R. Henderson
Senior Author

3.1 Introduction

There are many incentives to define lists of genes that are either essential for bacterial survival or important for normal rates of growth. Essential genes of bacterial pathogens may encode novel biochemical pathways or potential targets for antibacterial drug development. Disruption of genes required for rapid growth results in strains handicapped for exploitation in biotechnology. Conversely, growth of mutants defective in genes previously expected to be essential could reveal unexpected parallel biochemical pathways that can fulfil the essential function.

Multiple attempts have been made to generate definitive lists of essential genes, but there are still many discrepancies between different studies even for the model bacterium *E. coli* K-12. Two general approaches have been used to identify essential genes: targeted deletion of individual genes, as in the Keio collection of mutants (Baba *et al.* 2006), and random mutagenesis (Yamazaki *et al.* 2008; Nguyen & Valdivia 2012). Data from studies using different mutagenesis strategies have yielded inconsistent data and therefore conflicting conclusions. The aim of this chapter was to identify the essential genes of *E. coli* K-12 strain BW25113 using TraDIS. BW25113 is a well-studied model organism for which a complete gene deletion library is available (Baba *et al.* 2006).

A confounding factor in determining the 'essentiality' of a gene is the definition of an essential gene. Complete deletion of an essential gene results, by definition, in a strain that cannot be isolated following growth. However, it is well known that certain genes are required for growth under specific environmental and nutritional conditions (Tong *et al.* 2004; Nichols *et al.* 2011). Such genes can be considered conditionally essential. For the purposes of this study a gene is defined as essential if the

transposon insertion data reveal that the protein coding sequence (CDS), or a portion of the CDS, is required for growth under the conditions tested here. To minimise false positive classification of essential genes, an existing statistical model presented by Langridge *et al.* (2009) was adapted to include corrections for both gene length and genome length.

An additional challenge with defining essentiality in high throughput studies is an over-reliance on automated analysis of the data. For example, a consequence of relying only on quantification of the number of unique insertions within a gene is that genes with essential regions will be missed. If only part of a gene encodes the essential function, it should be possible to isolate viable mutants with transposon insertions in non-essential regions of the coding sequence (Yamazaki *et al.* 2008). Conversely, reliance on statistical analysis alone can also lead to over-estimation of the number of essential genes. This is a common result from insertion sequencing analysis (Grenov & Gerdes 2008). A low number of transposon insertion events within a gene, which fall below the statistical cut-off threshold, can be due to inaccessibility of the gene to transposition because of extreme DNA structure, exclusion by DNA-binding proteins, polarity effects due to insertion in a gene upstream of a co-transcribed essential gene, and location of the gene close to the replication terminus (Manna *et al.* 2007). The most frequent reason for a low number of insertions is that the product of the disrupted gene is required for normal rates of growth under the conditions tested. To minimize the possibility of incorrectly designating genes as essential or contributing to fitness, statistical analysis was supported with a gene-by-gene inspection of the insertion distribution within each individual gene.

3.2 Results and Discussion

3.2.1 Sequencing of a mini-Tn5 transposon insertion library

An amended TraDIS method was used to obtain data for a transposon mutant library of *E. coli* K-12 BW25113 (Langridge *et al*. 2009; Phan *et al*. 2013). The strain BW25113 was chosen because it is the parent strain for the Keio collection of deletion mutants and ideal for a direct comparison between datasets. The transposon library was constructed by a collaborator, Keith Turner, by transformation of a mini-Tn*5* transposon with a chloramphenicol resistance cassette into competent cells before growth overnight on selective medium. Individual colonies were pooled to construct the initial library, estimated to consist of approximately 3.7 million mutants. An Illumina MiSeq was used to obtain TraDIS data from two independent DNA extracts of the transposon library (TL), designated TL1 and TL2 (data collected by previous PhD student Ash Robinson). Raw data were checked for the presence of an inline index barcode to identify independently processed samples. This resulted in 4,818,864 sequence reads from TL1 and 6,189,409 from TL2. After verification of the presence of a transposon sequence and removal of poor quality data or short sequence reads, 3,891,339 (80.75%) and 4,387,970 (70.89%) sequence reads, respectively, were mapped successfully to the *E. coli* K-12 BW25113 genome (CP009273.1). There was a high correlation coefficient of 0.96 between the samples (Fig. 3.1A). The data were therefore combined to give a total of 8,279,317 sequences that were mapped to 901,383 unique insertion sites throughout the genome. Of the 901,383 identified insertion sites, 199,557 were represented by a single read. Similar numbers of insertions, 481,360 and 480,072, were found for both orientations of the transposon. The high density of unique insertion sites resulted in an average of 1 insertion every

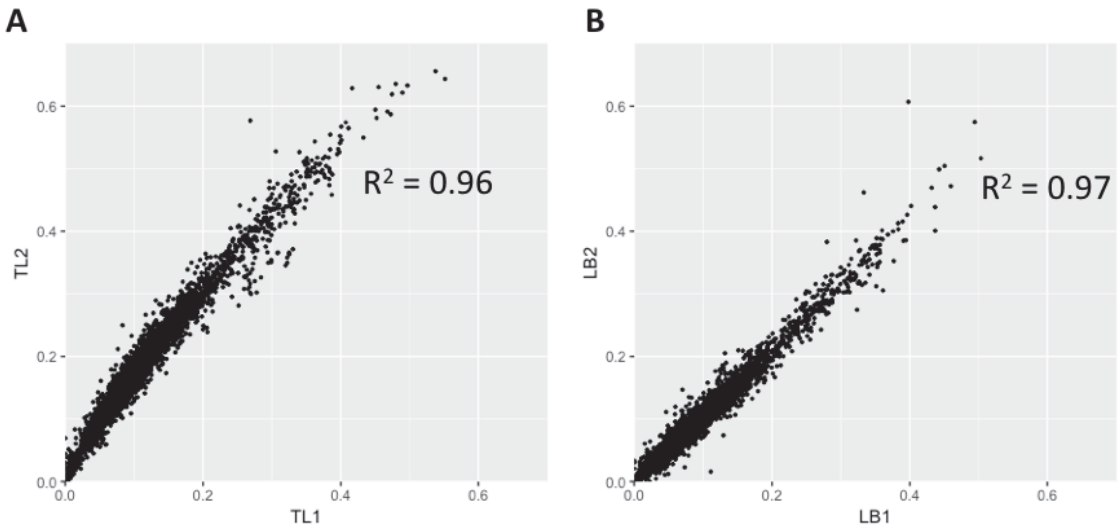

Figure 3.1 Sequencing of two independent replicates

(A) The correlation coefficient of gene insertion index scores for two sequenced technical replicates of the input transposon library (TL1 and TL2) (B) and following growth in LB (LB1 and LB2).

5.14 bp and a median distance between insertions of 3 bp. The transposon library was also grown in LB broth for ~5 generations to an OD_{600} of 1.00 and harvested and processed as above. There was also a high correlation coefficient between independent replicates LB1 and LB2 (Fig. 3.1B), and the data were pooled, resulting in a total of 10,584,188 mapped reads for this dataset.

3.2.2 Identification of putative essential genes by TraDIS

To determine whether a gene was essential or non-essential the numbers of insertions per CDS were quantified. CDS is defined as the protein coding sequence of a gene, inclusive of the start and stop codons. To normalize for gene length the number of unique insertion points within the CDS was divided by the CDS length in bases. This value was termed the insertion index score (IIS) and has been used previously as a measure of essentiality (Langridge *et al*. 2009; Phan *et al*. 2013; Curtis & Brun 2014; Christen *et al*. 2014). This metric is suitable given a sufficiently dense library (Solaimanpour *et al*. 2015).

The frequency distribution of the IISs was bimodal (Fig. 3.2), as previously shown by others (Langridge *et al*. 2009). Genes in the left-hand mode, which have a low number of transposon insertions, are assumed to be either essential for survival or genes that confer a very severe fitness cost when disrupted. The second mode contains genes with considerably more insertions; these are deemed non-essential (Fig. 3.2). Based on inspection of the distributions, an exponential distribution model was fitted to the mode that includes essential genes and a gamma distribution model was fitted to the non-essential mode. For a given IIS, the probability of belonging to each mode was calculated and the ratio of these values was termed the log-likelihood

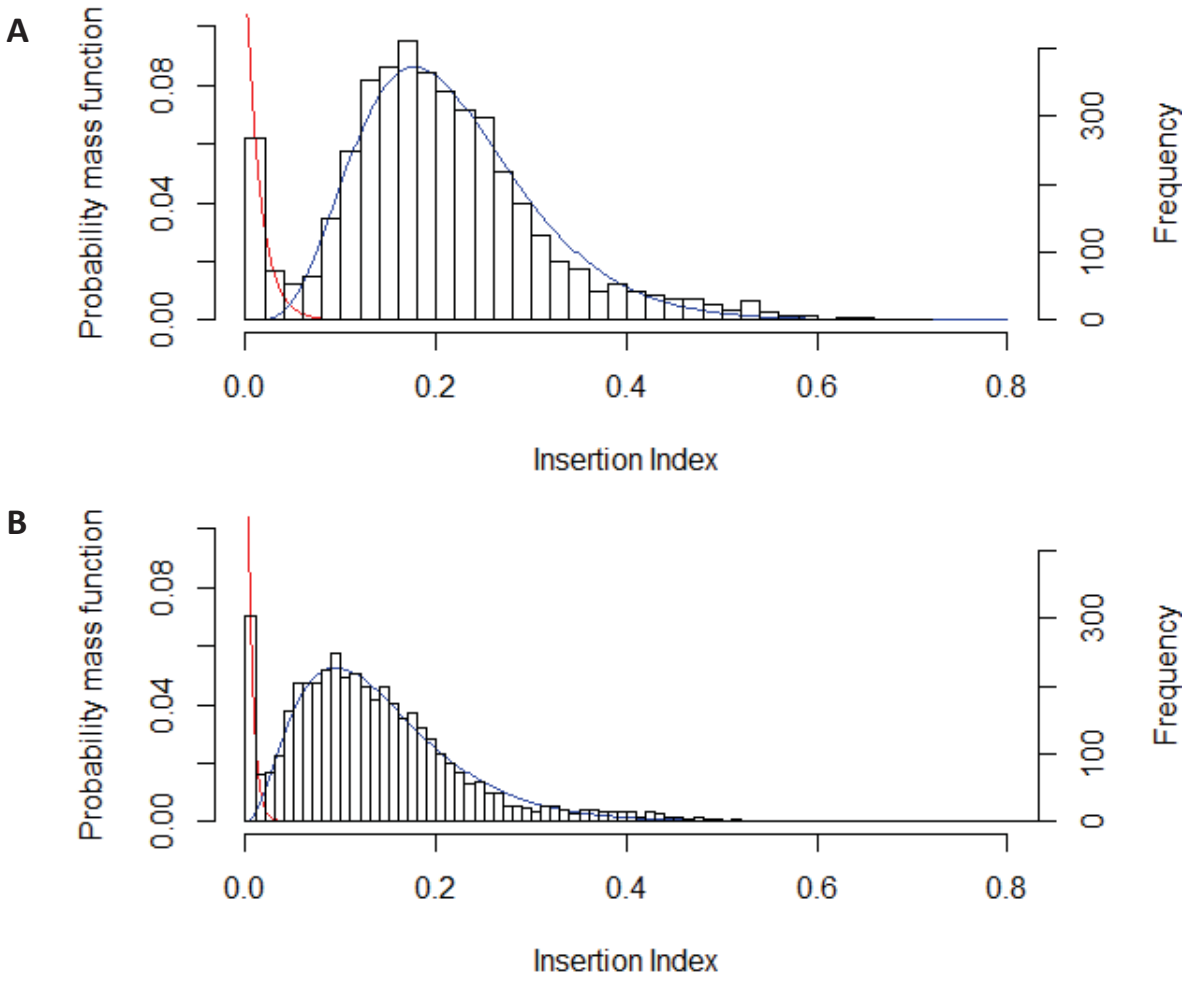

Figure 3.2 Distribution of insertion index scores

The insertion index score for each coding sequence was calculated as the number of insertions per CDS divided by the CDS length in bp to normalize for gene length. The frequency of insertion index scores was plotted for both TL (A) and LB data (B) and both followed a bimodal distribution. An exponential distribution model was fitted to the left mode that includes essential genes (red) and a gamma distribution model was fitted to the right, non-essential mode (blue). For a given insertion index score, the probability of belonging to each mode was calculated and the ratio of these values termed the log-likelihood score. A gene was classified as essential if its log-likelihood score was less than $\log_2(12)$ and was therefore 12 times more likely to belong to the red mode than the blue mode.

score. A gene was classified as essential if its log-likelihood score was less than $\log_2(12)$ and was therefore 12 times more likely to belong to the essential mode than the non-essential mode (S. Table 3.1). Using this approach, sufficient insertions were found in 3793 genes for them to be classed as non-essential, 162 genes were situated between the two modes and classed as 'unclear' and 358 genes in the mutant library were identified as essential (S. Table 3.1).

3.2.3 Quantification of transposon insertion density data around the genome

Transposon insertion density is known to be biased by proximity to the origin of replication (Chao *et al.* 2016). To understand how this impacts our data, the IIS of each CDS were plotted in order of annotation for the input TL dataset (Fig. 3.3). The IISs follow a sine wave-like pattern with a trough around CDS 1600 and a peak around CDS 3700, corresponding with the genome terminus and origin of replication, respectively. This is a common artefact of the library construction method and has been reported previously (Chao *et al.* 2013). The relative decrease in IIS around the terminus and the relative increase in IIS around the origin is due to 'gene dosage'. Under conditions of fast growth, chromosome replication can be initiated before termination of the previous replication event has been completed, resulting in a ratio of origin to terminus of 4:1. As a consequence, genes closer to the origin are present at a higher copy number than those near the terminus, and therefore have 2-4 times the chance of being disrupted by a transposon (Chandler & Pritchard 1975).

To account for this bias the insertion density was calculated within given macrodomains that encompass the origin and terminus respectively. In checking for uniformity of insertion density across genomic regions, the density of insertions

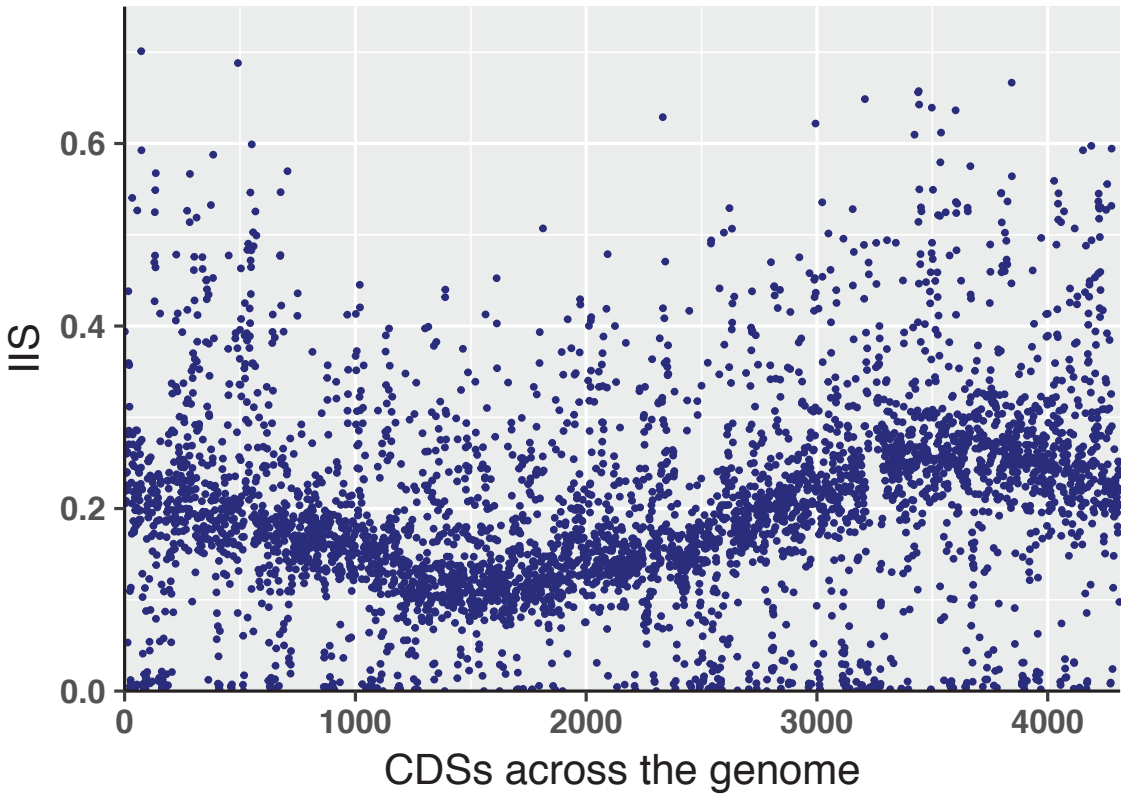

Figure 3.3 The effect of genomic position on insertion index score

The insertion index scores of the transposon library were ranked in annotation order (1-4313) and plotted. The insertion index scores vary according to genomic position, relative to the terminus and origin of replication.

around the terminus (taken as a region centred around *terABCD*) was slightly lower than the genomic average (a density of 0.142 in the surrounding 500 kb region, or 0.145 in the surrounding 1 Mb region, compared to a 0.195 average). This density change marginally increases the detection of false-positive essential genes in the vicinity of the terminus, but still represents an unprecedented level of coverage. As such, the data were not manipulated further, although it is possible to normalise for the positional bias mathematically (Chao *et al.* 2013).

3.2.4 Essential genes of the BW25113 genome

There are reports in the literature that, for some organisms, there is a higher proportion of essential genes on the leading strand than the lagging strand (Rocha & Danchin 2003). Therefore, the essential genes identified by TraDIS were annotated on a closed genome map (green) together with the transposon insertion frequency (blue), to visualise the genomic location of these genes (Fig. 3.4A). The essential genes of BW25113 identified by TraDIS are distributed evenly throughout the genome and on both strands (green, Fig. 3.4A). However, counter clockwise from the origin at ~3,700,000 there is a stretch of ~200,000 bp where essential genes are absent from the leading strand, presumably to minimise premature transcript termination as a result of DNA polymerase and RNA polymerase clashing (Srivatsan *et al.* 2010; Pomerantz & O'Donnell 2010; Couturier & Rocha 2006).

Although the increased density of transposon insertions around the origin does influence the IIS (section 3.2.3, Fig. 3.3), the relative difference between the IISs of essential and non-essential genes is great enough to differentiate between the two populations (as observed in the bimodal distribution of IISs). It is therefore assumed

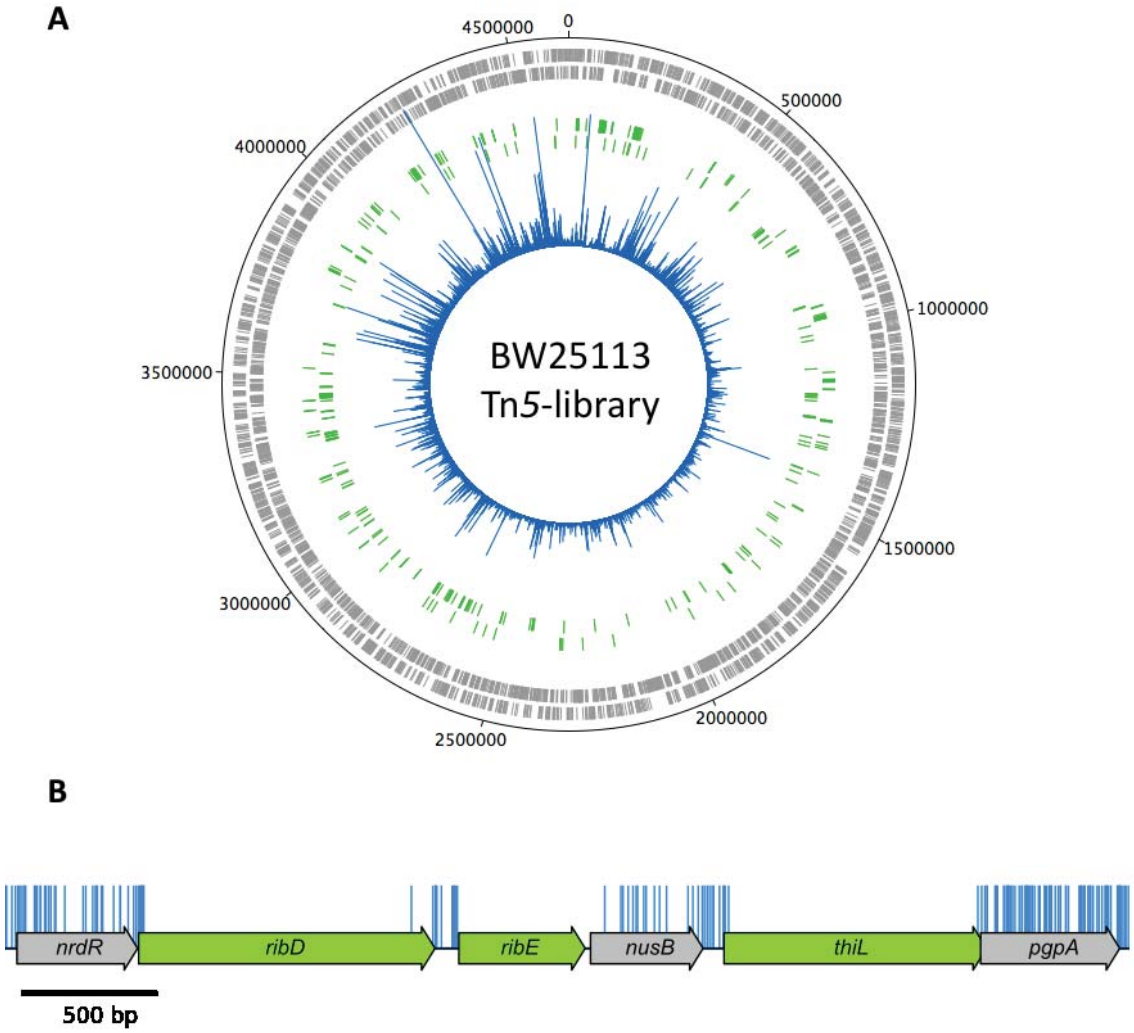

Figure 3.4 Transposon insertion sites within the BW25113 genome

(A) The frequency and location of transposon junction sequences from a mini-Tn*5* transposon library mapped to the BW25113 genome (CP009273.1). The outer track marks the BW25113 genome in bp starting at the annotation origin. The next two inner tracks correspond to sense and antisense CDS respectively (grey), followed by two inner tracks depicting the essential genes identified by TraDIS on the sense and antisense strands, respectively (green). The innermost circle (blue) corresponds to the frequency and location of recorded transposon insertion sites. (B) Representation of transposon insertion points across a portion of the BW25113 genome (blue), showing essential genes (green) and non-essential genes (grey). Blue bars correspond with transposon insertion sites along the genome and have been capped at a frequency of 1.

that the library is sufficiently dense to distinguish essential genes from non-essential genes. Representative examples of essential and non-essential genes are shown side by side in Fig. 3.4B.

3.2.5 Comparison between datasets of essential genes

The 358 putative essential genes identified in the TL data were compared to the essential genes as defined by the Keio collection and the Profiling of the *E. coli* Chromosome (PEC) database (Baba *et al.* 2006; Yamazaki *et al.* 2008). This comparison revealed 248 genes (59.5%) that were common to all three datasets (Fig. 3.5; Appendix 3.1). This agreement between all three datasets strongly supports the hypothesis that these genes are essential so they were not investigated further. An additional 169 genes were identified as potentially essential in only one or two of the datasets. These include 16 genes in the Keio and PEC lists of essential genes that were not identified by our analysis, 25 exclusive to Keio dataset, 18 exclusive to the PEC dataset, and 11 and 18 that overlapped between our method and the Keio or PEC datasets, respectively (Fig. 3.5). However, the largest subcategory of 81 genes is unique to our dataset.

3.2.6 Statistical analysis of the transposon insertion density data

Over-estimation of the number of genes that are essential has been noted in studies using transposon insertion sequencing (Grenov & Gerdes 2008). Conversely, failure of IIS analysis to recognise certain genes as essential has also been reported (Phan *et al.* 2013). Therefore, in conjunction with quantification of the frequency of insertions within a CDS, the data were analysed to determine the probability that a given stretch

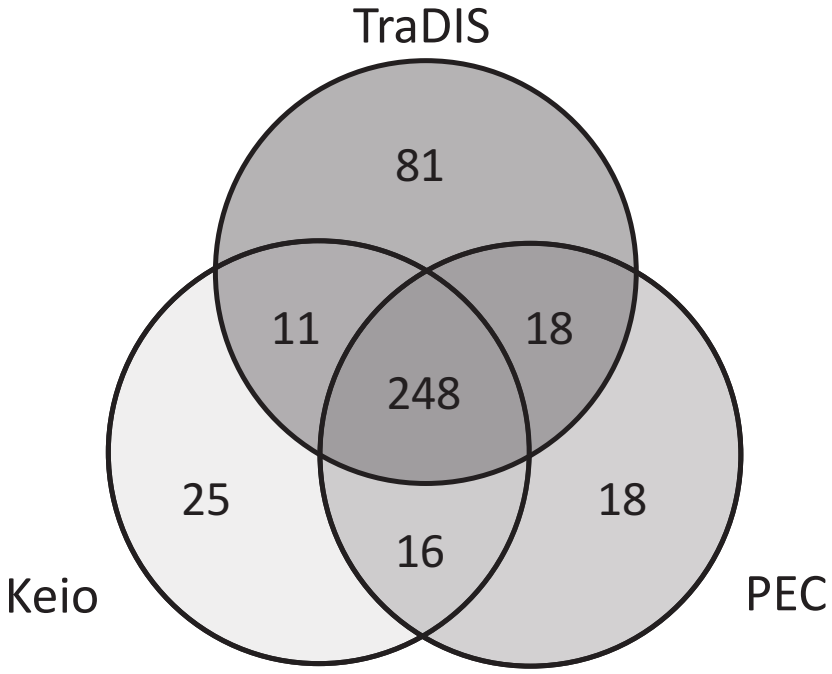

Figure 3.5 Comparison of essential genes between datasets

Candidate essential gene identified using TraDIS were compared to existing essential gene data. A 3-way comparison between the Keio collection of single gene knockouts, the Profiling of the *E. coli* Chromosome (PEC) online database and our transposon-insertion sequencing data identified 248 essential genes that were common to all three data.

of nucleotides without transposon insertion (an insertion-free region, IFR) occurred by chance. This work was done by Iain Johnston to refine existing published statistical analysis, and was used to inform whether a given IFR is likely to be essential. In short, the previously used Poissonian model $e^{(-f/x)}$ was replaced by a geometric model, whereby the probability of seeing k-1 "failures" (no insertion) followed by a "success" (an insertion) in a string of independent trials is $P(k)=\rho(1-\rho)^{(k)}$, where ρ is the probability of an insertion. Therefore, for an IFR of length L, the P value is $P = \sum_{k=L}^{\infty} P(k)$. This equation was then used to compute the P value of an IFR within a simulated library of random insertions of a given length (4.6 Mb) and density, corrected for gene length and genome size. For a genome of 4.6 Mb the probability of seeing one or more IFRs of at least length 75 bp is 0.05, and the corrected probability of one or more IFRs of 36 bp within a gene of 1,000 bp is 0.05. Similarly, for ρ_{gene} of 0.005, $L \cong 47$ bp. These values were used in subsequent analysis as the threshold for a statistically significant insertion-free region within the data. This represents the first study with a confident and genome-wide corrected detection resolution closest yet to approaching the length of the smallest annotated gene in our reference genome (CP009273.1), which is 45 bp.

3.2.7 Resolution of conflicts between datasets

A critical requirement for the validation of a list of essential genes is to explain why the statistical analysis of transposon insertion data failed to identify genes that the Keio library of deletion mutants and the PEC database identified as essential. Statistical analysis and manual inspection of the data were coupled with literature searches to rationalize conflicting results. The overall causes of discrepancy are

shown in Table 3.1 (individual genes are presented in Appendix 3.2). The two largest categories: "genes containing a transposon free region" and "errors in library construction", were divided further into sub-classes. Ultimately there are a wide range of explanations for discrepancies between datasets, but many of these can be explained by different methodologies used, definitions of the term 'essential' and statistical approaches (Fig. 3.6). Specific examples are discussed further below.

3.2.7.1 Genes containing transposon free regions

Manual inspection of the data revealed genes with transposon free regions that were large enough to be identified as significant using the equation defined in the previous section. These IFRs do not necessarily report that a gene is essential, but that the insertions within these genes are sufficiently sparse that the IFR is unlikely to have occurred by chance. These genes fall loosely into two groups. The first group contains genes for which the 5′ regions are essential and contain no insertions. However, there are transposon insertions in the non-essential regions of these genes, such as *ftsK* (Fig. 3.7A). FtsK is involved in correct segregation of the chromosome during division (Liu *et al.* 1998; Dubarry *et al.* 2010); the N-terminal domain of FtsK contains 4 transmembrane passes and is required for localization of FtsK to the septum (Draper *et al.* 1998; Yu *et al.* 1998; Dorazi & Dewar 2000). Consistent with the data, there is substantial body of literature reporting the essential function of the N-terminal domain (Wang & Lutkenhaus 1998; Draper *et al.* 1998; Dubarry *et al.* 2010). This is a common observation within transposon insertion data and arises when only the function of the N-terminus of the protein is essential (Christen *et al.* 2014). Initial analysis of transposon insertion data would lead to these genes being

Table 3.1 Causes of discrepancies between datasets

Cause	Number	
Transposon free region	**16**	
Essential domain[a]		10
Insertion free region[b]		6
Polar insertions	**5**	
Antitoxin	**3**	
Conditionally essential genes	**4**	
Essential after outgrowth		2
Nutrient-dependent essentiality		2
Errors in library construction	**35**	
Duplicated in the Keio collection		13
Slow growing mutant represented by few insertions		10
Keio deletion failure		9
Rapidly accumulates suppressor mutation		1
Not considered by our analysis	**14**	
Unknown	**11**	

[a] Defined as a gene with an IFR at one end of the CDS. [b] Defined as a gene with an IFR within the CDS framed by insertions either side.

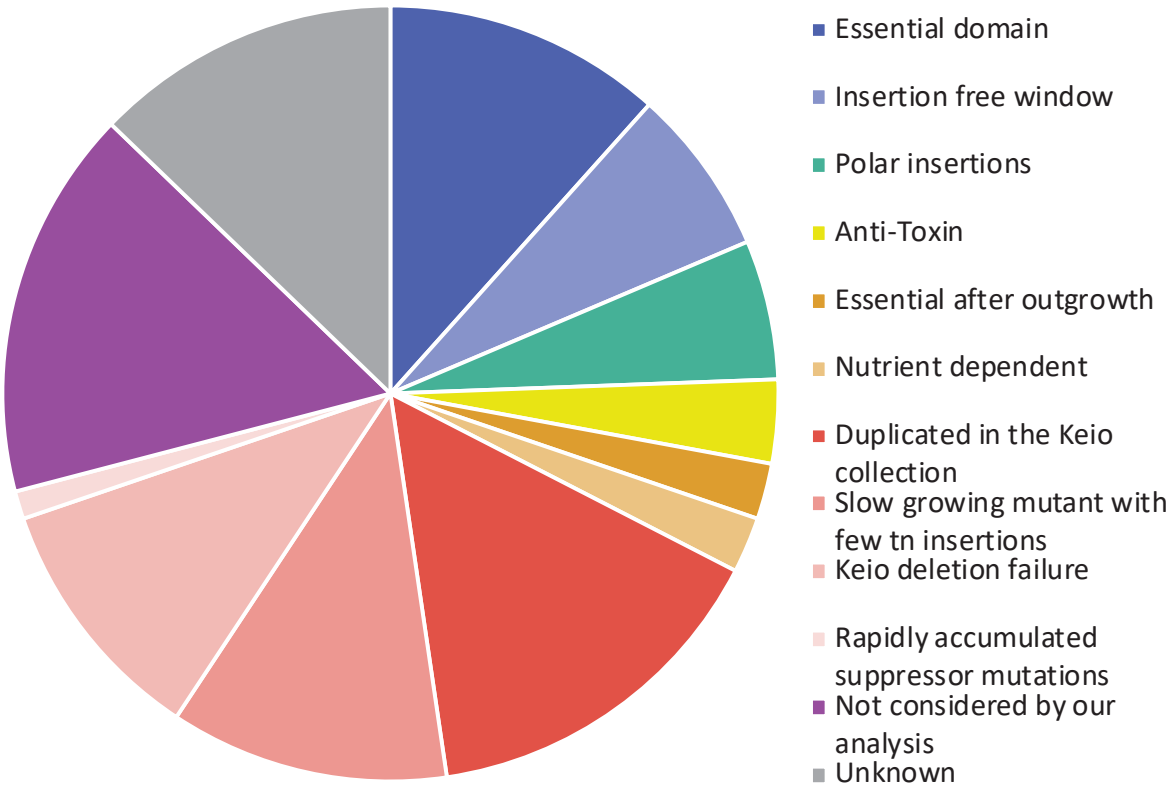

Figure 3.6 Source of discrepancies between datasets

The outlying genes of the Venn diagram, excluding those unique to our TraDIS dataset, were inspected to understand the source of discrepancy between datasets. Genes were grouped into the overarching categories of "Transposon free region" (blue), "Polar insertions" (green), "Anti-toxin" (yellow), "Conditionally essential" (orange), and "Errors in library construction" (red) with subcategories depicted in different shades of the colour. Genes not included in our analysis or that remain unclear are shown in purple or grey respectively.

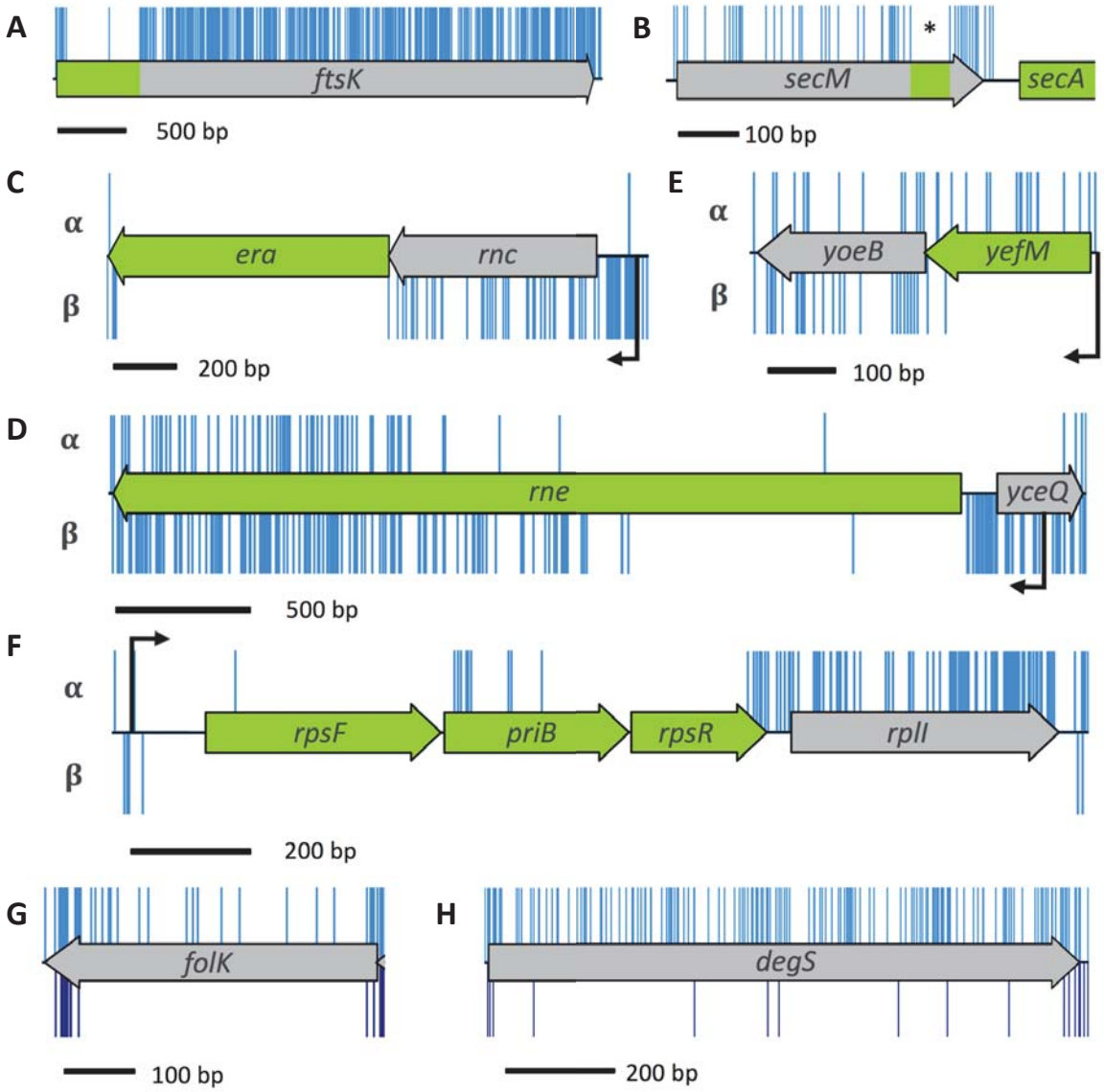

Figure 3.7 Insertion profiles of the discrepant genes between datasets

(A) Only a portion of the *ftsK* gene is essential for function. Such genes have a high IIS and consequently can be overlooked by automated statistical analysis. (B) *secM* contains a significant window (*) of 66 bp in which there were no transposon insertions. (C-F) Genes with transposon insertions in only one orientation. The α- and β-orientation of the transposon is depicted above and below the CDS track respectively, native promoters are shown in black. (G and H) Many transposon insertions were found along the full length of *folK* and *degS* (blue, upper track), however, most of these mutants were lost during outgrowth (dark blue, lower track).

incorrectly classified as non-essential, but attempts to construct a deletion mutant would fail. Indeed, previous transposon sequencing experiments failed to identify the essential nature of some of these genes when relying on statistical analysis alone (Phan *et al.* 2013).

The second group contains genes with transposon insertion sites throughout the CDS but which have an IFR that passes the significance threshold for essentiality. For example, there is a small 66 bp IFR within the coding sequence of *secM* (Fig. 3.7B). The *secM* gene is located upstream of the essential gene *secA*. These genes are co-transcribed and also co-translated, and *secM* is known to contain a translational stop sequence that interacts with the ribosomal exit tunnel to halt translation, acting as a translational regulator for *secA*. Specific mutations within the translational stop sequence are lethal unless *secA* is complemented by expression from a plasmid (Murakami *et al.* 2004). The dependence of *secA* translation on the *secM* CDS would explain the Keio classification as 'essential'. However, the IFR within *secM* does not fully correspond with the translation stop sequence, suggesting that there is more to be elucidated about the post-transcriptional regulation of these two proteins.

Other researchers have used different approaches to minimize false classification of essential genes during statistical analysis of the insertion profiles by applying a sliding window, quantifying the mean distance between insertions per gene, or variations of truncating the CDS, such as excluding the 3′ end, analysing only the first 60% of the CDS or analysing the central 60% of the CDS (Curtis & Brun 2014; Solaimanpour *et al.* 2015; DeJesus & Ioerger 2013; Freed *et al.* 2016; Zomer *et al.* 2012; Sarmiento *et al.* 2013; Yanjia J Zhang *et al.* 2012). However, window analysis

might overlook genes such as *secM* and analysing only the first 60% or the central 60% of the CDS would overlook genes such as *ftsK*.

In a sufficiently dense library the algorithmic approach used here is an appropriate method for identifying essential chromosomal regions. However, there are a number of IFRs within non-essential genes that are greater than 47 bp throughout the genome, which suggests that the null model of random insertions is not capturing the full structural detail of transposon insertion propensity. This suggests our modelling approach is not based on a perfect representation of biological reality and needs further refinement.

3.2.7.2 Insertion bias due to polarity effects

A common feature when creating insertion mutants is the introduction of off-target polar effects where expression of adjacent genes can be disrupted by the insertion (Fig. 3.8A). To mitigate against such polar effects the cassette was designed, by Discuva, to enable both transcriptional and translational read-through in one direction only. To confirm that transcriptional and translational read-through emanates from the transposon, the transposon was cloned in each orientation and for all three reading frames upstream of the *lacZ* gene in the transcription and translation expression vectors pRW224 and pRW225, derivatives of pRW50 (Lodge *et al.* 1992; Islam *et al.* 2012). Vector pRW224 retains an RBS for *lacZ* but no promoter, vector pRW225 has no promotor or RBS upstream of *lacZ*. Transcriptional read-through was confirmed for one orientation of the transposon, consistent with transcriptional read-through from the chloramphenicol resistance cassette into the downstream

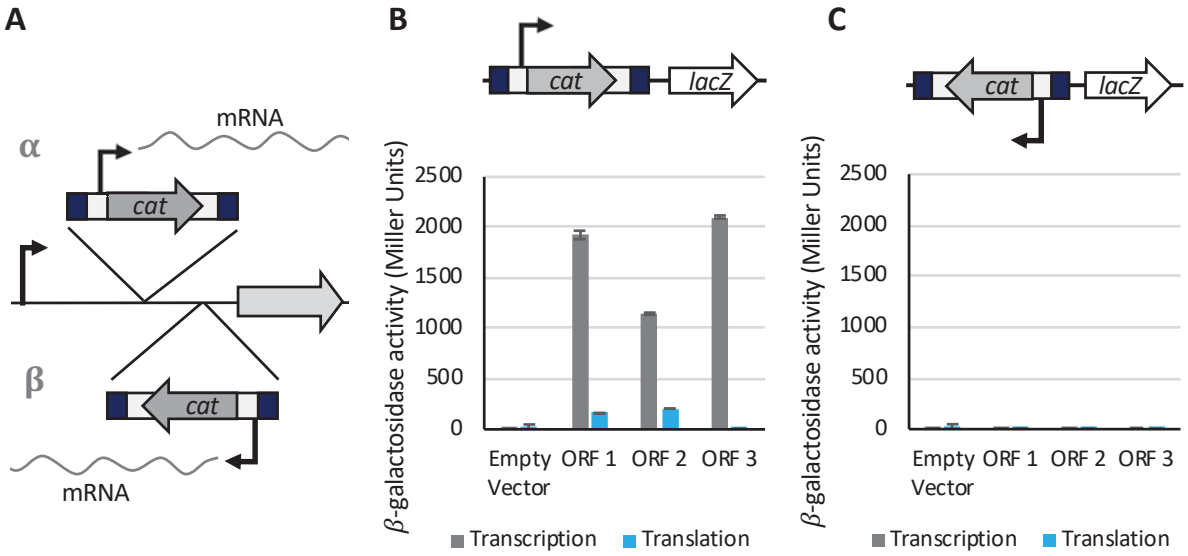

Figure 3.8 Transcription and translation initiation from within the transposon

Schematic representing the orientation of transposon insertions. The α-orientation of the transposon (upper track) corresponds with the chloramphenicol cassette oriented as depicted. The β-orientation (lower track) corresponds with transposon insertions in the opposite direction. (B and C) The full length mini-Tn*5* transposon was cloned into expression vectors pRW224 and pRW225 upstream of the *lacZ* gene, in each orientation, for all three open reading frames (ORF). Vectors pRW224 and pRW225 can be used to detect transcriptional and translational activity respectively. β-galactosidase activity was measured for 3 technical replicates; the mean value was plotted with error bars showing the standard deviation between replicates. (B) Transcriptional read-through was confirmed for one orientation of the transposon, consistent with the orientation of the chloramphenicol gene. Translational read-through from the mini-Tn*5* transposon was confirmed for 2 out of 3 open reading frames, consistent with GUG (ORF1) and AUG (ORF2) start codons in the transposon inverted repeat. (C) No transcriptional or translational read-through was detected for the opposite orientation of the transposon.

disrupted CDS (Fig. 3.8B and C). Translational read-through was identified for 2 of the 3 open reading frames that coincided with AUG and GUG start codons in the inverted repeat at the end of the transposon (Fig. 3.8B). More β-galactosidase activity was obtained from the construct in which the AUG codon was in frame than when the GUG codon that was in frame, confirming that translation initiation was stronger from the AUG codon. Therefore, transcription is initiated from within the transposon and translation is initiated from within the inverted repeat in one orientation only. This allows transcription and translation of downstream essential regions, even from within a CDS, depending on the orientation of the transposon. Such events can be identified by determining the DNA strand to which the sequencing data maps (Fig. 3.8A).

Analysis of the data reveals a number of chromosomal regions with insertions in only one orientation. Such insertion profiles can offer insight into transcriptional regulation of genes when considered in conjunction with neighbouring genes. For example, the gene *rnc* is located in an operon upstream of the essential gene, *era*. Only mutants with transposons that maintain downstream transcription of *era* are viable (Fig. 3.7C). Baba *et al*. categorized *rnc* as essential (Baba *et al*. 2006). However, in the case of the Keio library, construction of an *rnc* deletion mutant would disrupt the ability of the native promoter to drive downstream expression of the essential *era* gene, resulting in apparent lethality. Similarly, in both the Keio and PEC databases *yceQ* is listed as essential but many insertions were observed in *yceQ*, but in only one orientation (Fig. 3.7D). The gene is located upstream of the essential gene *rne* and is divergently transcribed. The promoter for *rne* is positioned within *yceQ* (Claverie-Martin *et al*. 1991; Ow *et al*. 2002) and deletion of *yceQ* would remove the

promoter for *rne* resulting in an apparent lethal effect. Our data reveal that while *era* and *rne* are essential, *rnc* and *yceQ* are not essential.

Similarly, several of the antitoxin genes are reported to be essential in the Keio library but not in our dataset or the PEC database. Antitoxins are required only if the corresponding toxin gene is functional. An example of this is *yefM,* within our data a substantial number of insertions were observed in one orientation only. Unlike *rnc* and *yceQ* where insertions maintained downstream expression, in the case of *yefM* the opposite is true; insertions that disrupt expression of the antitoxin but maintain downstream expression of the downstream toxin (*yoeB*) are lethal (Fig. 3.7E). Scrutiny of our data in this manner reveals that these genes are essential.

Another example of insertion bias is observed in a number of genes at the 3´ end of a transcript, such as *rplI* (Fig. 3.7F). While *rplI* is not reported as essential, it is worth noting because although insertions are restricted exclusively to one orientation within this gene, this cannot be explained by the positional context between an essential gene and promoter. One possible explanation for this observation is that transcription promoted from the transposon produces an antisense RNA that inhibits expression of an essential gene. Insertion bias, irrespective of the underlying cause, can result in false classification of genes when quantifying insertion frequency as these genes have half as many insertions relative to the rest of the genome. As such, these genes might be missed when analysing data with autonomous statistical approaches.

3.2.7.3 The effect of growth conditions on gene essentiality

In addition to the scenarios listed above, certain genes present challenges for binary classification of essentiality. For example, the genes might code for a protein that is

essential at a specific phase of growth or for growth under certain environmental parameters such as temperature or nutrient availability. Our data reveal a few of these conditionally essential genes. For instance, the Keio and PEC databases list *folK* as essential whereas in this dataset multiple insertions were detected within *folK* (Fig. 3.7G). Loss of *folK* disrupts the ability of the bacterium to produce folate, which is an essential metabolite. However, supplementation of the medium with folate abrogates the requirement for folate biosynthesis. In addition to *folK,* the Keio and PEC databases report *degS* as essential. DegS is a serine protease required for the activation of sigma E. In our dataset *degS* has a high density of insertions throughout the CDS suggesting that *degS* is not essential for growth on solid media (Fig. 3.7H). Consistent with this, there is substantial literature showing that *degS* mutants can be isolated, but they either lyse in the stationary phase of growth or rapidly accumulate suppressor mutations (Ades *et al.* 1999; Alba *et al.* 2001; Bass *et al.* 1996; Waller & Sauer 1996).

The conditional essentiality of such mutants can be tested by growing the transposon library in liquid broth. One would expect mutants lacking *degS* will lyse and that *folK* mutants will be outcompeted as the limited folate available in the medium is depleted. To test these scenarios, two independent samples of the transposon library were grown in LB at 37°C for 5 to 6 generations to an OD_{600} of 1.0 and were then sequenced. These samples, LB1 and LB2, resulted in 5,908,163 and 6,403,324 sequences of which 5,201,711 (88.04%) and 5,382,477 (84.06%), respectively, were mapped to the BW25113 genome. Insertion index scores were calculated as before (S. Table 3.2). As there was a high correlation coefficient of 0.97 between the gene insertion index scores of each technical replicate (Fig. 3.1B), the

data were combined to give a pool of 10,584,188 sequences. Scrutiny of our data revealed substantially fewer *degS* and *folK* mutants after growth in LB, relative to the input TL, supporting our hypothesis that they are conditionally essential (Fig. 3.7G and H).

3.2.7.4 Misclassification of genes due to differences in library construction methods and analysis

The difficulty in classifying a gene as essential through deletion analysis is the dependence on a negative result to inform classification. Thus, failure to knock out the gene might result in the false classification of a gene as essential. For example, the Keio database originally reported that *mlaB* (*yrbB*), a component of the Mla system for maintenance of lipid asymmetry, is essential. However, our data demonstrate that *mlaB* is non-essential and this is supported by the literature (Malinverni & Silhavy 2009; Thong *et al.* 2016). Similar outcomes were observed for several other genes (Appendix 3.3). The reason why knockouts of these genes were not obtained in the construction of the Keio library is unknown.

In addition to the false positive outcomes described above, several instances of false negative results within the Keio library database were noted. For example, both our TraDIS data and the PEC database identified 18 genes as essential that are reported as non-essential in the original Keio database. Subsequently, Yamamoto *et al.* (2009) demonstrated that for 13 of these mutants the target gene was duplicated during construction of the Keio library resulting in a functional protein; these genes are likely essential. Another difficulty that arises when targeting essential genes for mutagenesis is the potential to select for mutants with a compensatory mutation

elsewhere in the genome. Our data revealed that *hda,* an inhibitor of DNA replication re-initiation, is an essential gene but it is classified as non-essential in the Keio database (Kato & Katayama 2001; Katayama *et al*. 1998). Since the initial description of the Keio library, *hda* has been reported to be essential but *hda* mutants rapidly accumulate suppressor mutations that restore viability (Kato & Katayama 2001; Riber *et al*. 2006; Camara *et al*. 2003). This could be an explanation for the observed essentiality of some genes in the TraDIS dataset that were described as non-essential by others. These effects may arise when creating TraDIS libraries but the effects are masked by the large number of mutants in the population.

Similarly, in the PEC library where insertion density is low, when relying on single insertion mutants to inform essentiality, essential genes with an insertion in a non-essential region will be falsely classified as non-essential. An example of this false negative classification in the PEC database is *tadA.* The TadA protein is a tRNA-specific deaminase. The essentiality of *tadA* is reported in the Keio database and our dataset, and is supported by the literature (Wolf *et al*. 2002). The PEC database reports a single insertion site within the extreme 3′ end of the *tadA* gene.

A range of underlying causes behind dataset discrepancies have been identified and highlight that there are numerous possible insertion profiles for an 'essential' gene. As such, it is important to note that no single statistical method, to our knowledge, would fully identify every essential gene within a TraDIS dataset and that manual inspection of data is crucial.

3.2.8 Genes identified as essential only by TraDIS

There are 81 genes identified as essential using insertion index score data, which are not reported as essential in the Keio or PEC databases (Table 3.2). These genes fall

Table 3.2. Genes with low insertion index scores identified as essential by TraDIS only

Gene	Function[a]	Comments
Genes with no insertions		
glyA	Serine hydroxymethyltransferase	
pheM	*pheST* operon leader peptide	
rimM	Ribosome maturation protein	
rplK	50S ribosomal subunit	Duplicated in Keio collection[b]
rplY	50S ribosomal subunit	Duplicated in Keio collection
rpmF	50S ribosomal subunit	
rpmI	50S ribosomal subunit	
rpsF	30S ribosomal subunit	
rpsO	30S ribosomal subunit	Duplicated in Keio collection
rpsT	30S ribosomal subunit	
rpsU	30S ribosomal subunit	Duplicated in Keio collection
thyA	Thymidylate synthase	
ttcC	Pseudogene	
ynbG	Small protein	
Genes with a low frequency of insertions		
aceF	Pyruvate dehydrogenase, E2 subunit	
cydB	Cytochrome bd-I terminal oxidase subunit II	
cydD	Glutathione/L-cysteine exporter	
cydX	Cytochrome bd-I terminal oxidase	
dapF	Diaminopimelate epimerase	
dcd	dCTP deaminase	
fabH	β-ketoacyl-ACP synthase III	
fdx	Reduced ferredoxin	
folP	Dihydropteroate synthase	Duplicated in Keio collection
guaA	GMP synthetase	
hemE	Uroporphyrinogen decarboxylase	Duplicated in Keio collection
higA	Antitoxin of the HigB-HigA toxin-antitoxin system	
hipB	HipB antitoxin and DNA-binding transcriptional repressor	
holD	DNA polymerase III, Ψ subunit	

hscA	Chaperone for [Fe-S] cluster biosynthesis	
ihfA	Integration host factor (IHF), α subunit	
iraM	Anti-adaptor protein iraM, inhibitor of σS proteolysis	
iscS	Cysteine desulfurase	
iscU	Scaffold protein for iron-sulfur cluster assembly	
lipA	Lipoyl synthase	
lpd	Lipoamide dehydrogenase	
lpxL	Lauroyl acyltransferase	
lysS	Lysine-tRNA ligase	
mnmA	tRNA-specific 2-thiouridylase	
pdxH	pyridoxine 5'-phosphate oxidase / pyridoxamine 5'-phosphate oxidase	
priB	Primosomal replication protein N	Duplicated in Keio collection.
ptsI	PTS enzyme I	
rbfA	30S ribosome binding factor	
relB	RelB-antitoxin and DNA-binding transcriptional repressor	
rluD	23S rRNA pseudouridine synthase	
rnt	RNase T	
rpe	Ribulose-5-phosphate 3-epimerase	
rplA	50S ribosomal subunit	
safA	EvgS/EvgA and PhoQ/PhoP connector	
sucA	2-oxoglutarate decarboxylase	
sucB	Dihidrolipoyltransuccinylase	
tktA	Transketolase I	
tonB	TonB energy transducing system - TonB subunit	
trpL	*trp* operon leader peptide	
tusE	Sulfur transfer protein	
ubiE	Bifunctional 2-octaprenyl-6-methoxy-1,4-benzoquinone methylase and S-adenosylmethionine:2-DMK methyltransferase	
ubiG	Bifunctional 3-demethylubiquinone-8 3-O-Methyltransferase and 2-octaprenyl-6-hydroxyphenol methylase	
ubiH	2-octaprenyl-6-methoxyphenol hydroxylase	
ubiX	3-octaprenyl-4-hydroxybenzoate carboxy-lyase partner protein	
ybeY	Endoribonuclease involved in maturation of 16S rRNA and ribosome quality control	
ycaR	Conserved protein	
yciS	Lipopolysaccharide assembly protein	
ydaE	Rac prophage; zinc-binding protein	
ydaS	Rac prophage; predicted DNA-binding transcriptional regulator	
ydcD	Hypothetical protein	
yddL	Predicted lipoprotein	

ydfO	Qin prophage; predicted protein
ydhL	Conserved protein
yedN	Putative protein
yffS	CPZ-55 prophage
ygeF	Predicted protein
ygeG	Predicted chaperone
ygeN	Predicted protein
ygfZ	Folate-binding protein
yjbS	Hypothetical protein
ykfM	Hypothetical protein
ymfE	e14 prophage; predicted inner membrane protein
ymiB	Putative protein
yncH	Hypothetical protein
yobI	Small protein
yqcG	Cell envelope stress response protein
yqeL	Small protein

[a]Functions taken from Ecocyc: https://ecocyc.org/
[b]Yamamoto *et al.* 2009

into two groups, those with no insertions and the remainder with insertions in the CDS. The first group are most likely to be essential. For example, the S21 protein of the 30S subunit of the ribosome, *rpsU,* is essential in our data and has been described as essential by others (Fig. 3.9A; Bubunenko *et al*. 2007). However, in the Keio library there is a duplication event, which gives rise to a mutant that produces a functional protein (Yamamoto *et al*. 2009).

Scrutiny of our data for the remaining genes reveals there are additional essential genes with a low frequency of insertions. For instance, the DNA polymerase III ψ subunit, encoded by *holD*, has been described in the literature as an essential gene (Durand *et al*. 2016). The TraDIS data support that finding (Fig. 3.9B, Appendix 3.1). However, *holD* mutants are available in the Keio collection. The demonstration by Durand *et al.* (2016) and others (Viguera *et al*. 2003; Duigou *et al*. 2014) that *holD* mutants accumulate extragenic suppressor mutations at high frequency may explain why these mutants are considered non-essential in the Keio database and why a low frequency of insertions was observed in our TraDIS experiments.

A number of the essential genes unique to our analysis were not identified as essential in the Keio collection or PEC database simply because they are not included in either of these datasets. This is in part because the Keio collection of knockout mutants was based on available annotation data at the time (Riley *et al*. 2006). For example, the identification and location of *ynbG*, *yobI* and *yqcG* was published only in 2008 (Hemm *et al*. 2008). Our data show very sparse or no transposon disruption in these genes and consequently, these genes are potentially essential (Fig. 3.9C-E). Further validation studies would be required to confirm this.

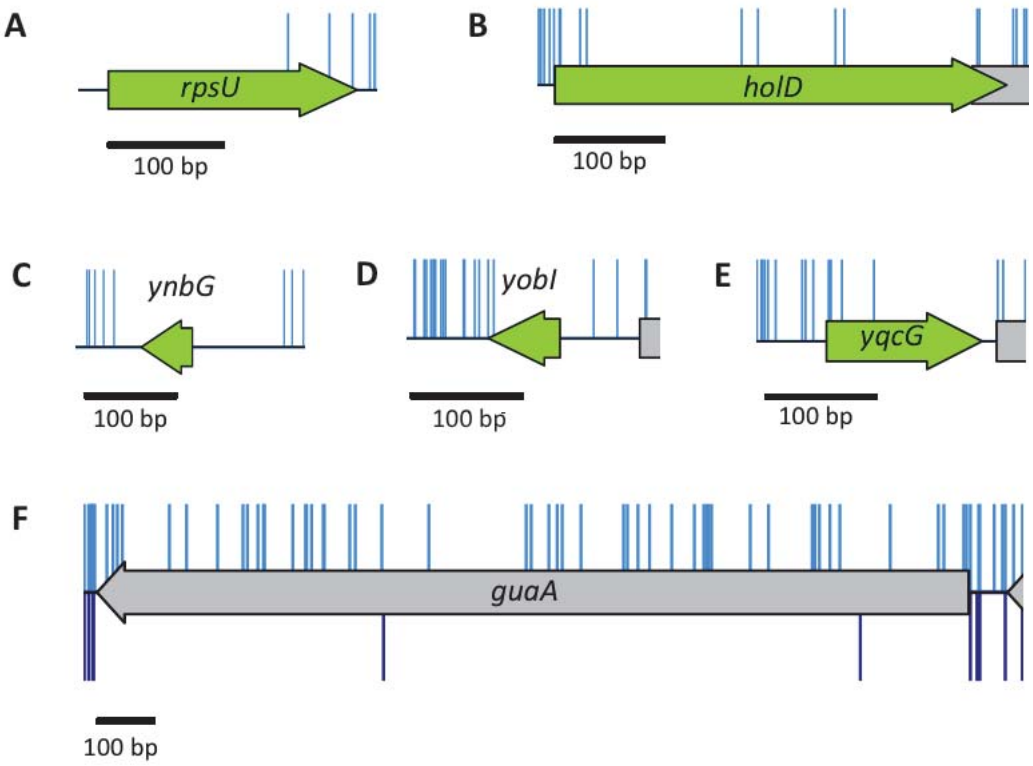

Figure 3.9 Essential genes unique to the TraDIS dataset

(A-E) There are very few or no insertions within these genes in the input TL (blue). (A & B) Low insertion frequency and literature support classification of these genes as essential. (C, D & E) Recently annotated genes with few or no insertions. The data suggest these are potentially essential or important for growth. (F) The gene *guaA* has a sufficiently low insertion index score to be classified as essential after initial statistical analysis (blue, above). Following outgrowth, there are few *guaA* mutants (dark blue, below), consistent with literature reports that *guaA* mutants have a growth defect.

As mentioned previously, over-reporting of essential genes may occur when non-essential genes have low insertion index scores. Such low insertion index scores may arise due to attenuated growth. An example of the mis-classification of a gene because transposon insertion results in a fitness cost and attenuated growth is *guaA*. The *guaA* gene encodes a GMP synthetase and a Δ*guaA* mutant is auxotrophic for guanine (Lambden & Drabble 1973). The low insertion index score results in *guaA* being classed as essential despite having many insertions. The fitness effect was confirmed by growing the library in LB, as such mutants are outcompeted (Fig. 3.9F), and the literature supports the fact that this gene is not essential and has an altered growth rate (Davies & Drabble 1996). The decrease in detectable *guaA* mutants following outgrowth is likely due to depletion of available guanine in the medium.

3.2.9 High resolution features within a TraDIS dataset

Manual inspection of a TraDIS dataset can reveal additional information that might go unnoticed in a high throughput analysis pipeline. A common observation from this and previous detailed analysis of data from saturated transposon libraries is the ability to determine, at the bp level of resolution, the boundaries of essential regions within a gene. An example of an essential gene with a dispensable 3′ end is *yejM* (*pbgA* in *Salmonella* Typhimurium), which is thought to code for an IM anchored cardiolipin transport protein. Only the 5′ end of the CDS is essential, up to and including codon 189, which corresponds with 5 transmembrane helices of the protein structure. The C terminus of the protein is a periplasmic domain that is dispensable for viability (Fig. 3.10A; De Lay & Cronan 2008; Daley *et al.* 2005; Dalebroux *et al.* 2015). Our TraDIS data revealed insertions in codons 186 and 189. Analysis of the transposon

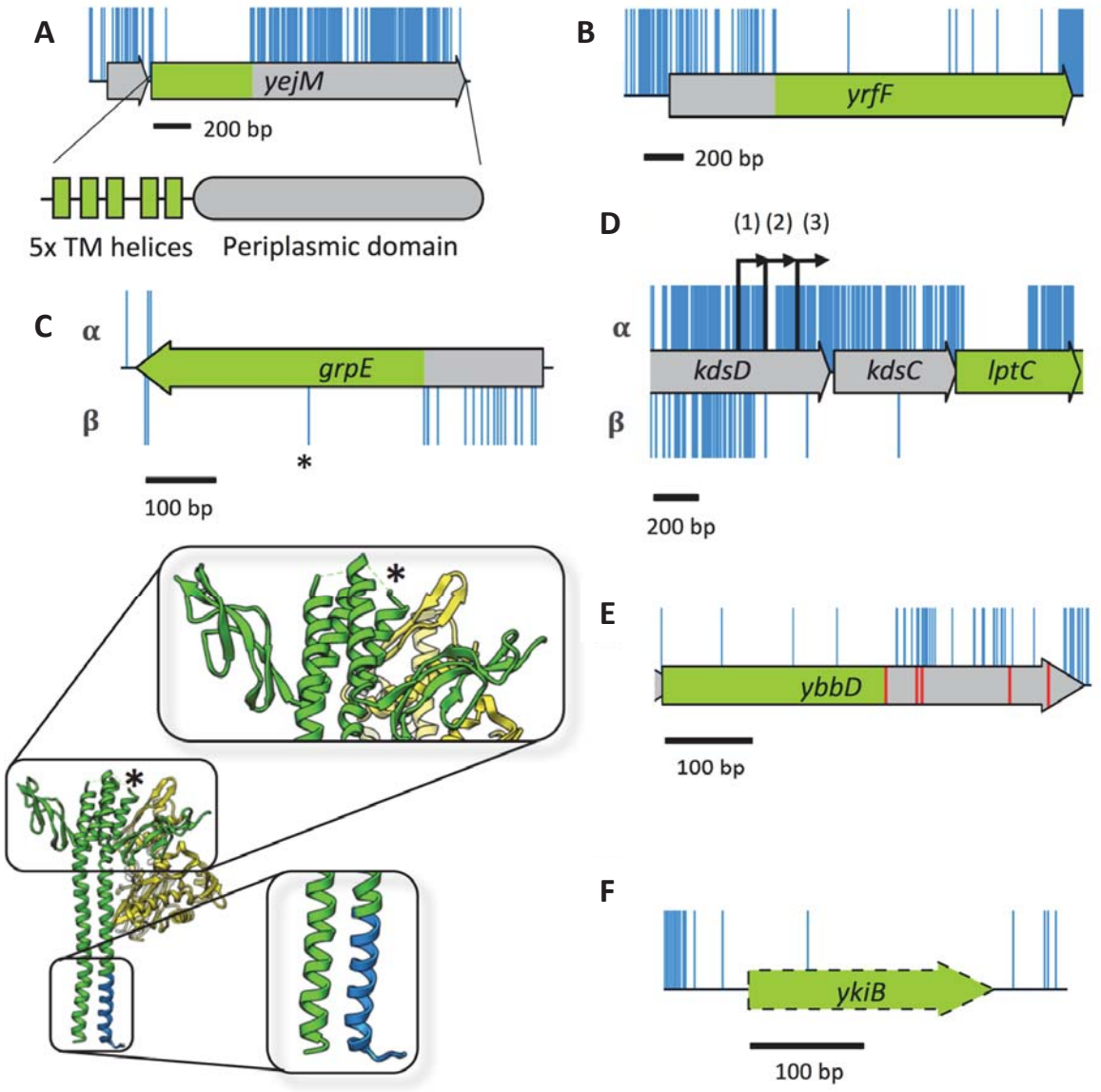

Figure 3.10 Additional features identified through detailed analysis of high-resolution insertion data

(A) Insertions within the CDS but not along the full length correspond with a non-essential periplasmic domain. The 5′ end of the CDS has no insertions and corresponds with 5 essential transmembrane domains of YejM. (B) Insertions within *yrfF* suggest a dispensable 5′ domain. (C) The *grpE* gene tolerates transposon insertions in the 5′ end of the CDS (blue), but only in the orientation that maintains expression of the downstream protein (lower track, β-orientation). The GrpE protein

forms a dimer (green) which interacts with DnaK (yellow). Transposon insertions in specific regions of the protein do not disrupt GrpE interaction with DnaK (blue). An additional, single, insertion point in the center of the *grpE* CDS (*) maps back to a turn between 2 helices of the GrpE protein. The data reveal dispensable sections of the GrpE protein and boundaries in secondary structure. (D) Insertions immediately upstream of *lptC* have an insertion orientation bias. Only insertions that maintain expression of *lptC* are tolerated within *kdsC* (α-orientation). The gene *lptC* has 3 promoters (‘1’, ‘2’ and ‘3’), the insertion boundary indicates that promter ‘2’ is the essential promoter. (E) Pseudogene *ybbD* contains many more insertions after the first stop codon (red), suggestive that the truncated CDS may still be functional and essential. (F) The pseudogene *ykiB* is not annotated in the BW25113 genome (CP009273.1) and has a single insertion within the CDS.

orientation at these points revealed that they corresponded with the same transposon insertion location but, due to the 9-bp duplication introduced by the transposon, in different transposon orientations. The introduced transposon sequence maintains codon 189, completely consistent with previously reported results (De Lay & Cronan 2008; Dalebroux *et al.* 2015).

In addition, as a result of the transposon design, a further feature of our TraDIS data is the identification of genes with dispensable 5′ ends. An example of this is *yrfF*, which encodes an inhibitor of the Rcs stress response (Fig. 3.10B; Cano *et al.* 2002; Cho *et al.* 2014). This phenomenon, while less well covered in the literature, is not surprising given that Zhang *et al.* report equal likelihood of a required intragenic region residing at the 5′ or 3′ end of a gene, albeit in *Mycobacterium tuberculosis* (Yanjia J Zhang *et al.* 2012). These mutants will only be viable if the remaining CDS can be translated into a functional product and one would expect to find an orientation bias where the transposon drives downstream transcription and translation of the essential region.

Interestingly, inspection of our data revealed essential genes with isolated insertions within the coding sequence. An example of this is *grpE*. The *grpE* gene codes for the essential nucleotide exchange factor that forms a dimer and interacts with the DnaK/J complex (Harrison *et al.* 1997). The insertion occurs only in the orientation that maintains expression of the remaining CDS (Fig. 3.10C). Mapping of the site of transposon insertion onto the previously determined protein structure of GrpE indicated the insertion occurred within part of the gene encoding a flexible linker between two α-helices (Fig. 3.10C). This suggests that similar to the GAL4 protein of *S. cerevisiae,* used in yeast two-hybrid system (Fields & Song 1989), the

adenylate cyclase from *Bordetella pertussis* used in the bacterial two-hybrid system (Karimova *et al.* 1998), and split GFP (GFP1-10 and GFP11; Cabantous & Waldo 2006), a functional GrpE can be expressed as two separate essential domains that interact to form a functional protein. Therefore, the data presented here can be exploited to identify every essential gene that could be used in a protein-fragment complementation assay to develop similar protein-protein interaction screens. Importantly, this TraDIS library has an unprecedented sub-CDS level of resolution that can demarcate changes in the protein secondary structure.

Another fine-resolution mapping feature of our transposon data is the identification of the promoter position for essential genes, as also previously reported by Christen *et al.* (2014). An example of this is the promoter for *lptC*, which is located within the *kdsD* gene. Altogether, there are three promoters within the *kdsD* gene (kdsCp3, kdsCp2 and kdsCp1; Martorana *et al.* 2011; Sperandeo *et al.* 2006). However, our data show that insertions that maintain expression of *lptC* are tolerated within the *kdsD* gene up to kdsCp2; insertions stop short just before the kdsCp2 -35 consensus sequence, excluding a single insertion between the -35 and -10, and a single insertion further downstream (Fig. 3.10D). These results suggest that kdsCp3 is dispensable and at least kdsCp2 is required for adequate expression of *lptC*. As in the case of *grpE*, this is another example of the unprecedented level of genetic detail that can be obtained from this high throughput method.

Finally, within our data a number of transposon-free sections that do not correspond with the annotated features of the genome were observed. This can occur when a start codon is mis-annotated (Christen *et al.* 2014), or translation might initiate at a secondary start codon downstream of the transposon insertions. Alternatively, a

pseudogene annotation may extend beyond the first stop codon. One example of this is *ybbD,* a pseudogene with a premature stop codon classified as non-essential in our dataset. However, the annotation of *ybbD* in the BW25113 genome extends beyond the first stop codon, whereas in other genomes it does not, and the region from the methionine translation start codon to the first stop codon passes the threshold for essentiality (Fig. 3.10E). This suggests that the truncated *ybbD* might be functional and also essential. In addition, a transposon free region corresponding to the gene *ykiB* was identified (Fig. 3.10F). This gene is annotated in W3110 but not in BW25113, despite the nucleotide sequence being identical. Our data suggest that these genes have a significant role in viability or growth but further investigation is required to test this. These examples highlight the importance of having a fully annotated and curated reference genome for mapping data. However, even the highly-studied K-12 genome with its wealth of annotation information retains some as yet unexplained transposon free regions. Thus, TraDIS can help identify regions of genomes where annotation is incorrect or incomplete.

3.2.10 Validation of gene essentiality

Within the 3-way comparison of essential gene lists, the second largest category was those identified only by TraDIS. Therefore, the literature was searched for information about these 81 genes that might explain the low insertion index scores. Most of the genes were reportedly not essential but contribute to fitness and have a marked growth-defect when disrupted or deleted. However, there were some genes for which there was no literature available, or reports were inconclusive or conflicting. Therefore, to resolve remaining uncertainties, 10 genes were selected for validation by

P1 transduction. A further 2 genes (*yddK* and *ypjC*), which were identified by eye and statistically deemed 'unclear' (Fig. 3.11 A & B), and 2 genes discussed previously (*ybbD* and *ykiB*) were also chosen for validation experiments. Strains from the Keio library of deletion mutants were used as donor strains and cassette-disruption mutants were constructed via P1 transduction as described in the materials and methods.

The P1 transduction was successful for all 14 of the candidates. Correct insertion of the kanamycin cassette was checked by PCR using oligonucleotides designed to flank the target gene. The PCR was successful for 13 of the primer pairs, and 12 of the candidates produced the expected band size for correct insertion of the kanamycin cassette (Fig. 3.11C). These genes are most likely not essential. The essentiality classification of genes *ydaE* and *guaA* remains inconclusive, as the PCR to check *ydaE* failed, and the PCR to check the *guaA* mutants (8 were tested, data not shown) produced no visible size change between the wildtype (WT) and the supposed mutant. In conclusion, no new essential genes were identified, but most remaining uncertainties were resolved.

3.2.11 Growth of a transposon library in LB

To identify genes important for growth under standard laboratory conditions, the transposon library was grown in LB for ~5-6 generations until an OD_{600} of 1.00 was reached. The IISs were calculated as before, and compared to the input library (Fig. 3.12A). There was a positive correlation between the two datasets; however, as expected, there was a visible overall decrease in the IISs of the LB dataset reflected by the downwards shift of the data from the diagonal mid-point of the graph (x=y). Specific outliers have been labelled, and their individual insertion profiles shown in

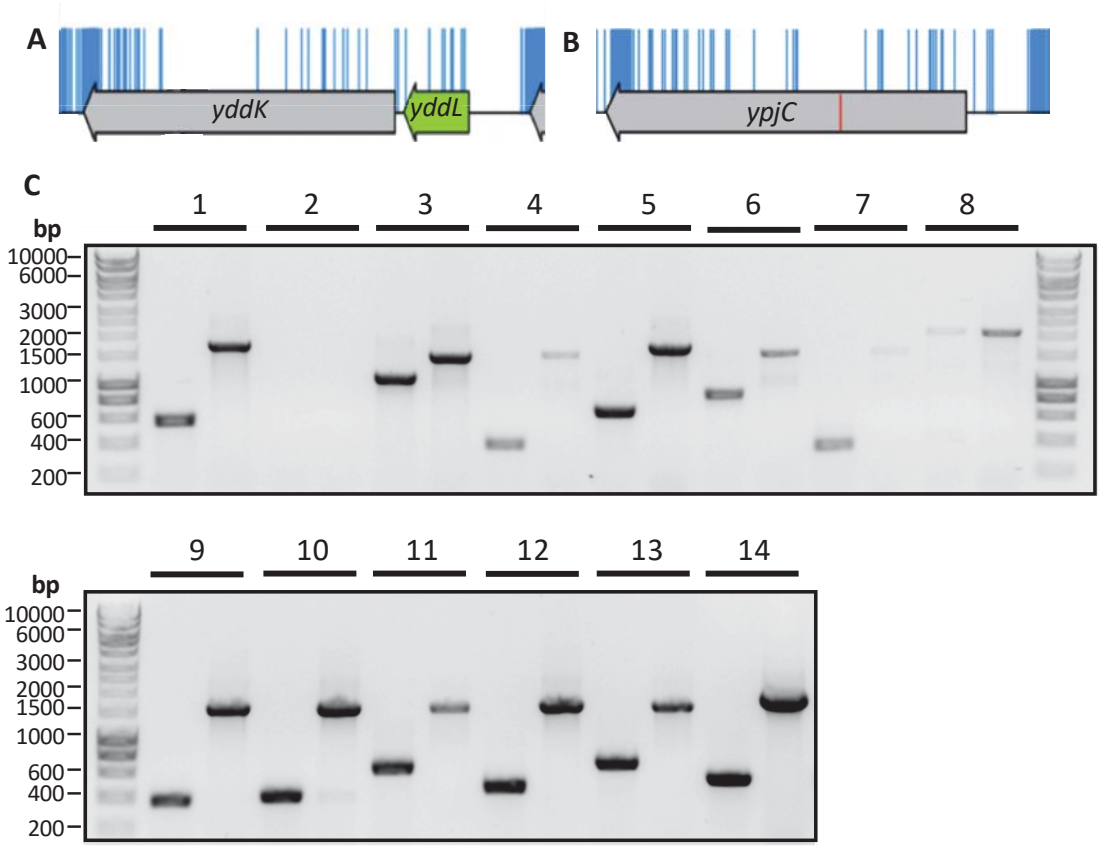

Figure 3.11 Validation of the gene essentiality of unresolved genes

(A and B) The insertion profiles of two genes identified by eye to have a low frequency of insertions within the CDS, but did not meet the statistical threshold to be classified as essential. (C) PCR amplification of genes targeted for deletion shown in pairs, with the WT on the left, and the corresponding p1 transductant on the right. The primers were designed to flank the following genes: (1) *ybbD*; (2) *ydaE*; (3) *yddK*; (4) *ydhL*; (5) *yedM*; (6) *ymfE*; (7) *yncH*; (8) *guaA*; (9) *rpmF*; (10) *rpmI*; (11) *ydcD*; (12) *ykiB*; (13) *ypjC*; and (14) *ybgT*.

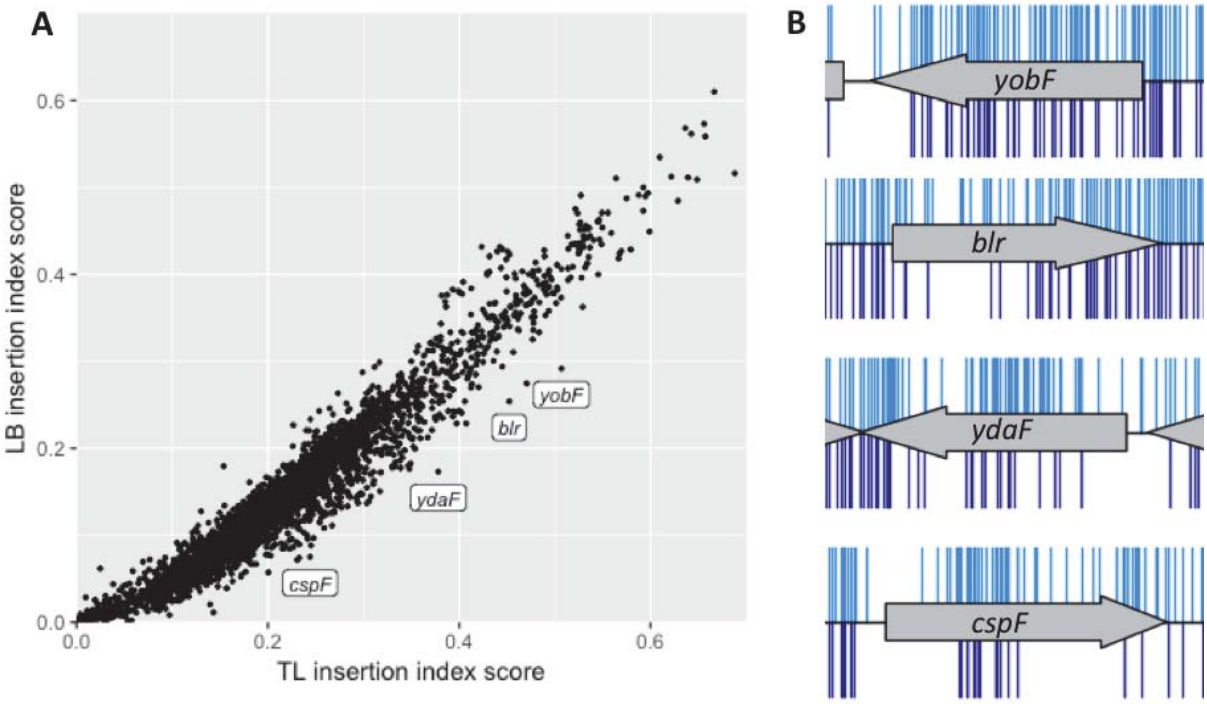

Figure 3.12 Comparison of insertion index scores between the input transposon library and mutants grown to an OD of 1.00

(A) The insertion index scores of genes within the LB outgrowth dataset (collected at an OD of 1.00) were compared to the insertion index scores of genes within the input transposon library (TL). 4 genes, *yobF*, *blr*, *ydaF* and *cspF*, with notably lower insertion index scores following outgrowth are labelled beneath the points. (B) The insertion profiles before and after outgrowth (light and dark blue, respectively) are shown above and below the gene tracts for all 4 genes.

Fig. 3.12B, with total insertions of the input pool above (blue), and LB outgrowth below (navy blue). The insertion profiles of some of these genes appear to show regions of sparse transposon insertion instead of a decrease in insertion density along the full length of the CDS. This could suggest conditionally-essential domains of these genes, which are required for growth in LB, or the IFRs could indicate an underlying essential sRNA within the annotated CDS.

Another feature that these genes have in common is that they are all shorter than average (with lengths of *yobF*, 126 bp; *blr*, 213 bp; *ydaF*, 156 bp; *cspF*, 144 bp). Perhaps these genes are outliers due to IIS bias; genes of a shorter length will be less well represented within the total mutant pool because there are fewer positions of the CDS that can be occupied by a transposon. Therefore, shorter genes will have fewer unique mutants that can be sampled and the effect of stochastic mutant loss will be more pronounced.

3.2.12 Growth kinetics of the outlying mutants

To determine whether the outlying mutants identified in the above section have a growth defect, gene-disruption mutants were grown in a 96-well plate in LB broth in triplicate and the OD_{600} was recorded every 30 min. Gene-disruption mutants were constructed by transduction of the kanamycin resistance cassette from the corresponding mutant of the Keio collection into BW25113. Growth defects of the mutants were marginal, no significant differences in growth rate were detected during exponential growth. However, the final OD_{600} recorded at 16 h for the *cspF*, *ydaF* and *yobF* mutants were significantly lower than WT ($P \leq 0.05$, Two-way ANOVA with Dunnett's multiple comparisons test), while the ODs measured during early stationary

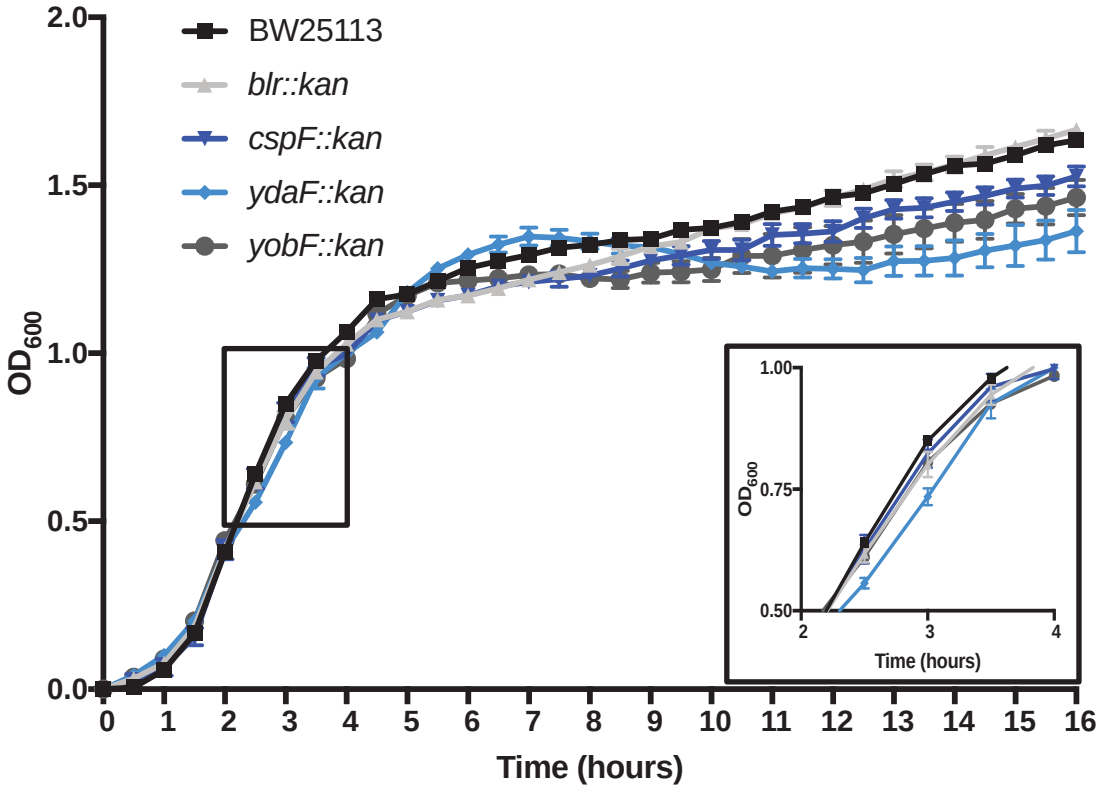

Figure 3.13 Growth kinetics of the targeted gene disruption mutants

Three biological replicates of strains were grown in a 96-well plate and the optical density recorded every 30 mins for 16 h. The mean OD_{600} was plotted and the standard deviation represented by error bars. The WT, BW25113, was included for comparison (black). A close-up of the growth profiles between 2 and 4 hours is shown in the inset.

phase were significantly lower than WT for the *blr* mutant ($P \leq 0.05$, Two-way ANOVA with Dunnett's multiple comparisons test). All 4 of the mutants reach an OD_{600} of 1.00 10-30 min later than the WT (3.5 h vs. 3.75-4 h; Fig. 3.13), which was also the OD_{600} at which the transposon mutant library was harvested. The marginal growth defect observed in a 96-well plate might be sufficient to account for the IIS decrease when the mutants are grown in competition. Badarinarayana *et al.* (2001) report even a 2% decrease in doubling time can result in 50% loss from the population over 30 generations if grown in competition. If these strains are slow growing mutants that are out-competed by a population of transposon mutants, it might be expected that if the transposon library was grown for longer in LB that these mutants would gradually disappear from the population.

3.2.13 Growth of a library of transposon mutants over time

To determine the effect of outgrowth on a library of transposon mutants, the library was grown again in LB, and samples were harvested at ODs of 1.00, 2.00, 3.00 and 4.00. The IISs were calculated for each gene, under each condition, and these were plotted in comparison to the input library (TL) insertion index scores (Fig. 3.14A; S. Table 3.3). Overall, as the OD_{600} increased, the IISs decreased. This was more obvious for genes with higher IISs, where the data were more sparse allowing a clearer view of the underlying trend. This result is not surprising and equates to the loss of mutants during outgrowth, either stochastically or through competition.

However, the loss of mutants upon entering stationary phase was not as pronounced as expected. While there was a trend of decreasing IIS with an increase in OD, the populations largely overlapped. For example, the IISs of the mutants

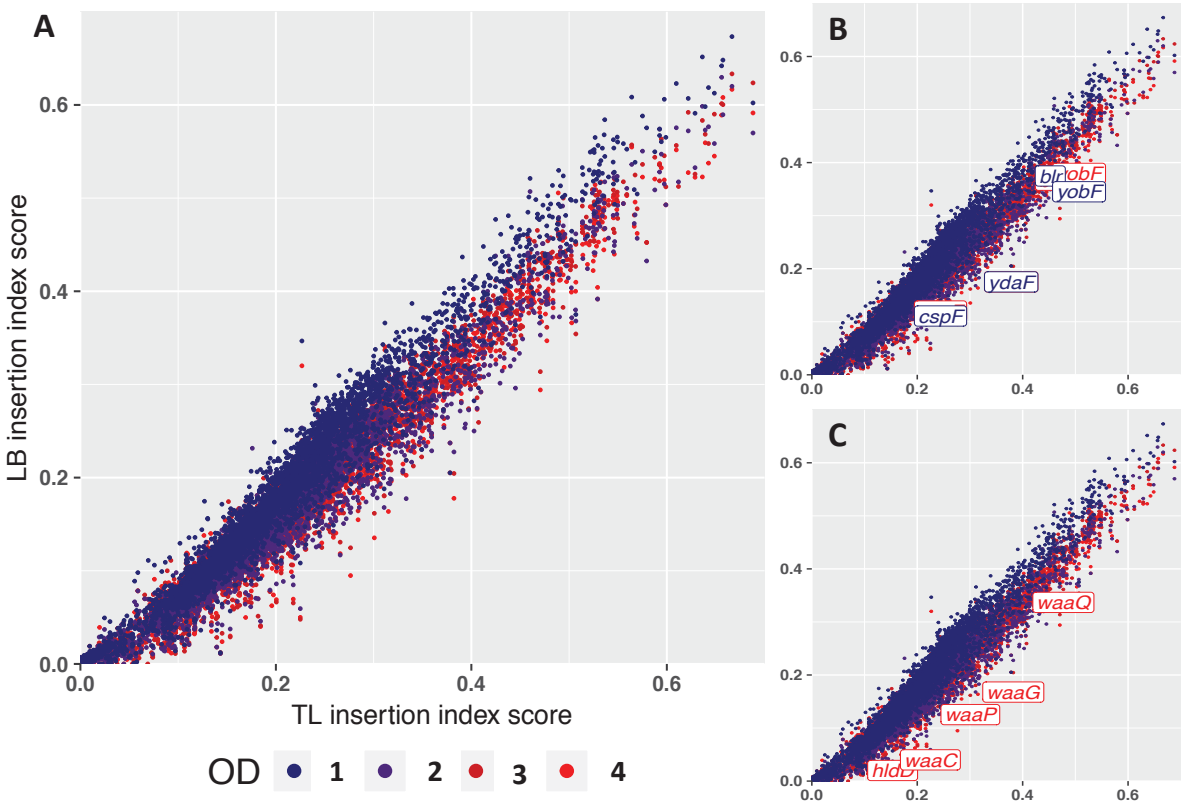

Figure 3.14 Growth of a transposon mutant library over time

(A) The transposon library was grown in LB from an OD_{600} of 0.05 and harvested at ODs of 1.00, 2.00, 3.00 and 4.00. The insertion index scores were calculated for every gene under each condition (OD=1.00, dark blue; OD=2.00, purple; OD=3.00, maroon; OD=4.00, red) and were compared to the input pool (TL). (B) Mutants identified in fig. 3.12 are indicated with labels at a fixed distance below the point. Labels are given only for OD=1.00 (blue) and OD=4.00 (red) datasets. (C) Genes of the *waa* operon drop below the general population trend following outgrowth to an OD of 4.00 in LB (red).

identified in 3.2.8 were expected to gradually decrease with outgrowth. When the IISs of these mutants are labelled for the OD_{600}=1.00 (navy blue) and OD_{600}=4.00 (red) datasets the IISs do not differ widely (Fig. 3.14B). Nevertheless, when the total data were inspected by eye within artemis, an obvious trend in decreasing IISs over outgrowth was observed for a specific subset of genes: the *waa* operon (Fig. 3.14C). When the difference between IISs of the TL and OD_{600}=4.00 datasets were calculated and ranked largest to smallest, 5 out of the top 20 genes belong to the *waa* operon. Furthermore, these 5 genes (*waaC*, *waaF*, *waaG*, *waaP* and *waaQ*) code for products that synthesise linkages within the inner-core of LPS. This could suggest that the LPS inner-core has an important structural role upon entering stationary phase, and disruption to these genes results in a decrease in fitness upon entering stationary phase growth in broth. It is possible that the depletion in IIS of these genes between an OD_{600} of 1.00 and an OD_{600} of 4.00 is due to lysis of the corresponding mutants. While these genes are not conditionally-essential, this result highlights the sensitivity of a method like TraDIS.

3.3 Conclusion

In summary, comparison of the TraDIS data with data from two previous studies of *E. coli* K-12 under standard laboratory conditions revealed 248 genes designated as essential in all three datasets. The findings of this chapter demonstrate the importance of visual analysis of TraDIS data to avoid automation bias in designating genes as essential or non-essential. In particular this chapter addresses the discrepancies between different conclusions drawn from transposon mutagenesis data and gene deletion studies. The most common causes of discrepancies are due to

combined essential and dispensable regions of a gene or polarity effects of neighbouring genes. Essential genes that contain both essential and non-essential regions will statistically appear non-essential if judged only on insertion index scores. While, polarity effects due to neighbouring essential genes in the same transcription unit can halve the number of possible viable mutants, which might influence statistical classification of gene essentiality. It remains to be seen whether one single statistical approach can capture all the target genes of interest in a screen for essential genes. This highlights a limitation of using 'disruption-frequency' as a measure of essentiality, a result emphasised by the finding that certain genes and pseudogenes with very low insertion index scores can be deleted and are therefore non-essential. Evidently there are alternate explanations for a low insertion frequency other than gene essentiality which remain to be explored.

In summary, TraDIS can not only be used to identify the essential genome of an organism but also for fine-resolution mapping of features across the genome. Importantly, TraDIS data are a valuable resource that can be re-inspected following the discovery of new features within a given genome. Finally, these data reveal that there is more to be understood about the genome structure and organization of the model organism *E. coli* K-12, which further coupling of modelling and experimental approaches will help to elucidate.

Chapter 4

Identification of New Genomic Features from TraDIS Data

4.1 Introduction

The high level of resolution offered from a dense transposon mutant library can aid the identification of additional genomic features. TraDIS data can reveal essential non-coding sRNAs, essential tRNAs, essential protein-binding sites and small protein coding sequences within an open reading frame that would otherwise go undetected by standard screening methods (Christen *et al.* 2014; Yanjia J. Zhang *et al.* 2012; Mann *et al.* 2012). By analysing the orientation of transposon insertions within an *E. coli* K-12 transposon library, new genes, the approximate location of the transcription start sites of essential genes, essential promoters, and antisense features can also be identified. In this chapter some distinctive insertion profiles are highlighted and used to discover previously unreported features of the *E. coli* genome.

The aim of this chapter was to interrogate the TraDIS data, which is the densest library published to date, to determine whether additional unannotated genes could be identified. This was achieved by defining transposon insertion profiles that might reveal novel unannotated features such as ORFs, transcripts or regulatory antisense RNAs. Recognition of similar insertion profiles in transposon mutagenesis datasets would assist genome annotation of less-well characterized genomes.

4.2 Results and Discussion

4.2.1 Identification of a recurring insertion profile within essential operons

The data were inspected by eye to establish whether any recurring themes or patterns within the TraDIS data could be identified. As previously mentioned in Chapter 3, orientation bias of transposon insertions can reveal co-transcription of non-essential and essential genes, as well as the position of a promoter for an essential gene or

operon. The insertion profile of co-transcribed essential genes has three distinguishing features: 1) no insertions are found within essential genes, 2) insertions in both orientations define the operon boundaries, and 3) insertions only in the orientation that maintain expression of downstream genes mark the boundaries between essential genes. A specific example of this profile was found for the three-gene *hemA* operon (Fig. 4.1). All three of the genes, *hemA*, *prfA* and *prmC*, encode essential products (Avissar & Beale 1989; Johnson *et al*. 2012). Using this model, it should be possible to recognise the presence of an unannotated essential gene within an essential operon.

4.2.2 Extrapolation of insertion data to reveal a putative unannotated gene in an essential operon

Having identified an insertion profile for operons of essential genes, the TraDIS data were revisited to determine whether any operons with essential genes might contain unannotated essential genes. One striking example is the *accBC* operon represented in figure 4.2. This operon is framed by transposon insertions in each orientation, essential genes *accB* and *accC* cannot be disrupted, and upstream of *accB* there are transposon insertions in only the α-orientation, but not for the full length of the untranslated region (UTR) between the promoter and *accB*. The presence of a putative start codon within the UTR suggests that there could be an additional essential CDS in this operon.

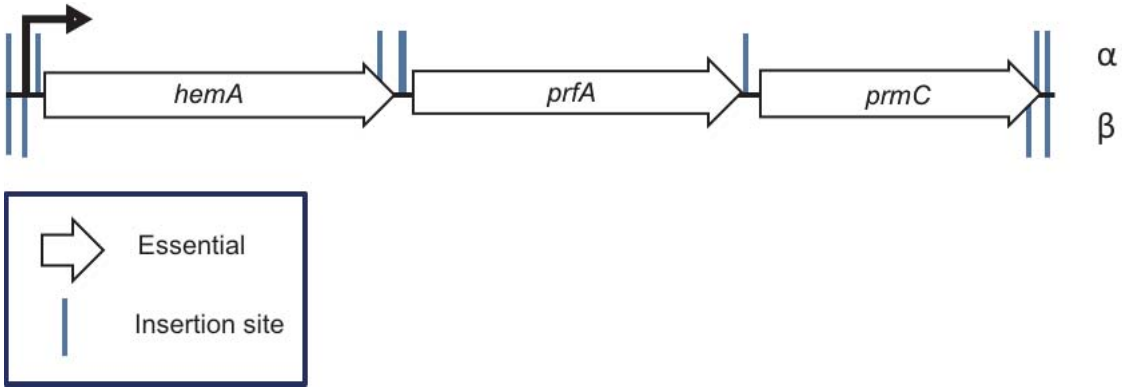

Figure 4.1 A common insertion profile within an operon of essential genes

A cartoon schematic representing transposon insertion sites surrounding essential genes within an operon. Transposon insertion sites are represented by blue bars above and below the gene track, corresponding with each orientation of the transposon. Transposon insertions above (α-orientation) correspond with transcription read-out from the transposon into the downstream essential genes. Transposon insertions below (β-orientation) correspond with transcription read-out from the transposon antisense to the annotated operon. Transposon insertions in both orientations frame the operon. Within the operon, insertions are only present in intragenic regions and only in the α-orientation that maintains expression of downstream essential genes.

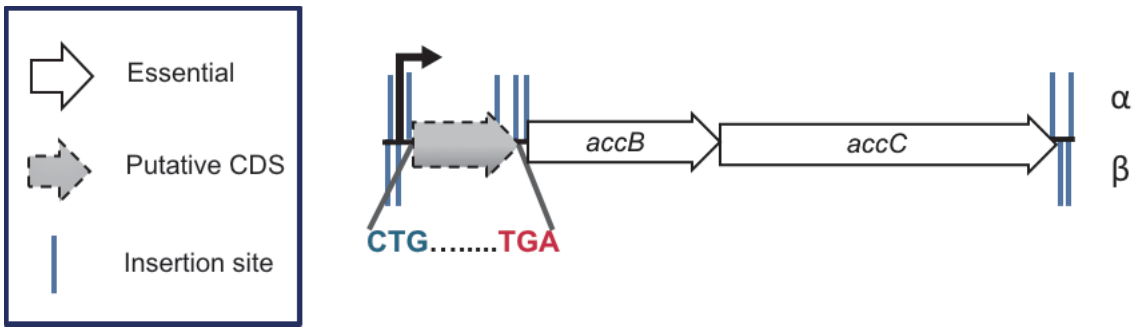

Figure 4.2 Schematic of insertion bias within the *accB* operon

Transposon insertion sites are represented by blue bars above and below the gene track, corresponding with each orientation of the transposon. Transposon insertions within the 296 nt 5′ UTR are restricted to one orientation, and do not span the full length of the UTR. Transposon insertion sites either side of the IFR may frame a previously unannotated essential gene in the *accBC* operon.

4.2.3 Characterisation of the transposon-free region upstream of the accB gene

To understand why there is an IFR within the *accB* operon, the insertion profile and operon structure were investigated further. The *accB* operon consists of two genes, *accB* and *accC*, which code for biotin carboxyl carrier protein (BCCP) and biotin carboxylase, respectively. The genes *accBC* coupled with *accA* and *accD* code for subunits of the acetyl-CoA carboxylase (ACC) complex, which converts acetyl-CoA to malonyl-CoA; the first committed step of fatty acid (FA) biosynthesis (James & Cronan 2004). The co-expression of *accBC* is reported to be highly co-conserved in a wide range of distinct bacterial species (James & Cronan 2004). As such, the regulation mechanism is likely to be highly conserved, but is yet to be fully characterised.

Transcription of this operon is initiated from two overlapping promoters, accBp1 and accBp2 (S. Li & Cronan 1993; My *et al.* 2015). The accBp2 promoter activity is stronger than accBp1; the accBp2 promoter is activated by FadR, however the FadR box is between the -35 and -10 sites of accBp1, and so FadR also represses accBp1 activity (My *et al.* 2015). FadR, fatty acid degradation regulon, is a transcription factor that negatively regulates fatty acid degradation and activates expression of unsaturated fatty acid biosynthesis. In the presence of long chain fatty acyl-CoA, FadR binding to the FadR box upstream of accBp2 is alleviated (My *et al.* 2015). The accBp1 promoter is repressed by BCCP (James & Cronan 2004). Both FadR and BCCP mediate transcription from upstream of the first transcription start site.

There is a 5′ UTR of 296 bases between the first transcription start site and the *accB* start codon. Transposon insertion data for this region show insertions within the UTR, but are restricted to one transposon orientation, consistent with maintaining

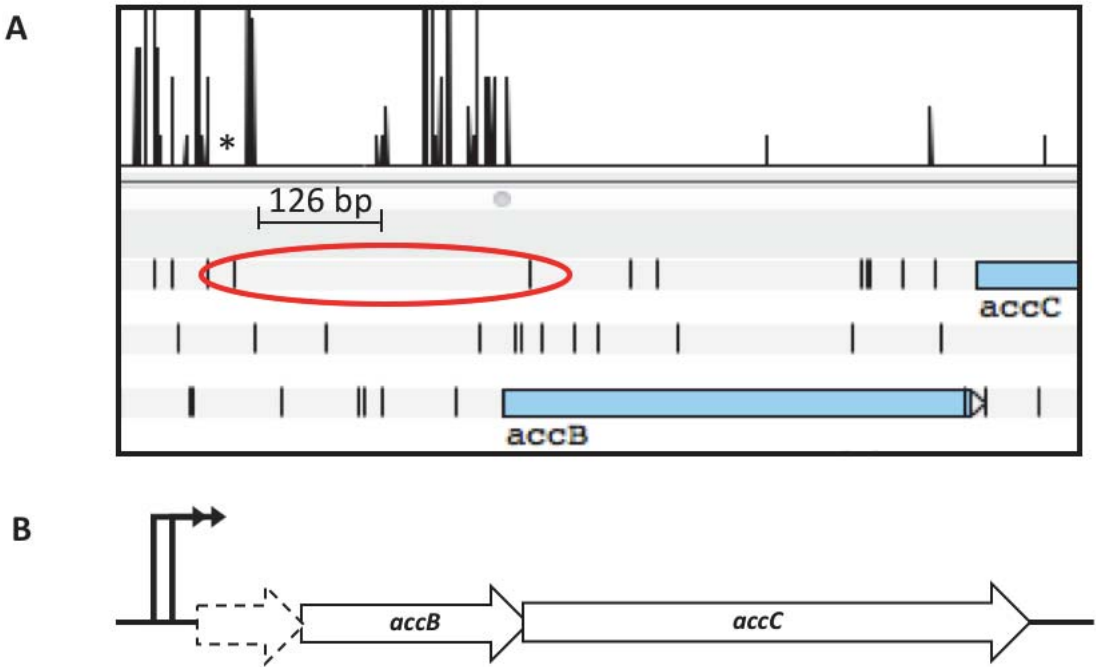

Figure 4.3 Characterisation of the *accB* 5′ UTR

(A) Transposon insertion-site data for the *accB* operon viewed within artemis. The top track of black peaks corresponds with the location and frequency of recorded insertion sites, and are capped at a frequency of 10. The lower 3 tracks represent the 3 reading frames of the sense strand, annotated genes are shown in blue, and stop codons are marked by black vertical bars. Between the *accB* promoters (*) and the *accB* start codon, there is a window of 126 bp with no insertions. The IFR coincides with a large ORF upstream of *accB* (red). (B) A schematic of the *accB* operon and promoters, with the putative CDS shown by a dashed arrow.

expression of *accB* and *accC* (Fig. 4.3A). However, within the UTR, there is an IFR of 126 bp between the promoters (shown by the asterisk in the figure) and *accB*, which is statistically significant for this transposon library (Fig. 4.3A). This would suggest that the IFR is functionally important, or is required for correct expression of the downstream essential genes. The IFR also overlaps a reported stretch of highly 'bent' DNA and a putative CDS (Li and Cronan 1993), within a prominent ORF (circled in red; Fig. 4.3B). Taken together, this information suggests there could be an additional essential gene within the *accB* operon, upstream of *accBC* (Fig. 4.3C).

To determine whether the orientation specific insertions within the *accB* 5′ UTR demarcate an additional essential CDS between *accB* and its promoters, the *accB* promoters and the putative start codons within the ORF were cloned into transcription and translation expression vectors pRW224 and pRW225 upstream of *lacZ* (Fig. 4.4A). Although detection of translation initiation is an indirect approach, this was initially favoured over insertion of a chromosomal myc tag due to the overlap of the putative CDS and the *accB* start codon. Secondly, insertion of a myc tag might disrupt any native regulatory control. Furthermore, the absence of any transposon insertions within this overlap region suggested chromosomal constructs would be challenging.

Four putative start codons were identified within the ORF and the corresponding promoter fragments were cloned in frame with *lacZ* into the pRW vectors. The *accB* start codon was also cloned into both vectors as a positive control. All ten constructs were confirmed by PCR (Fig. 4.4B) and Sanger sequencing. The β-galactosidase activity of 3 biological replicates was recorded following the protocol described in materials and methods. This was repeated on more than one occasion and a representative single dataset is presented showing the mean and standard

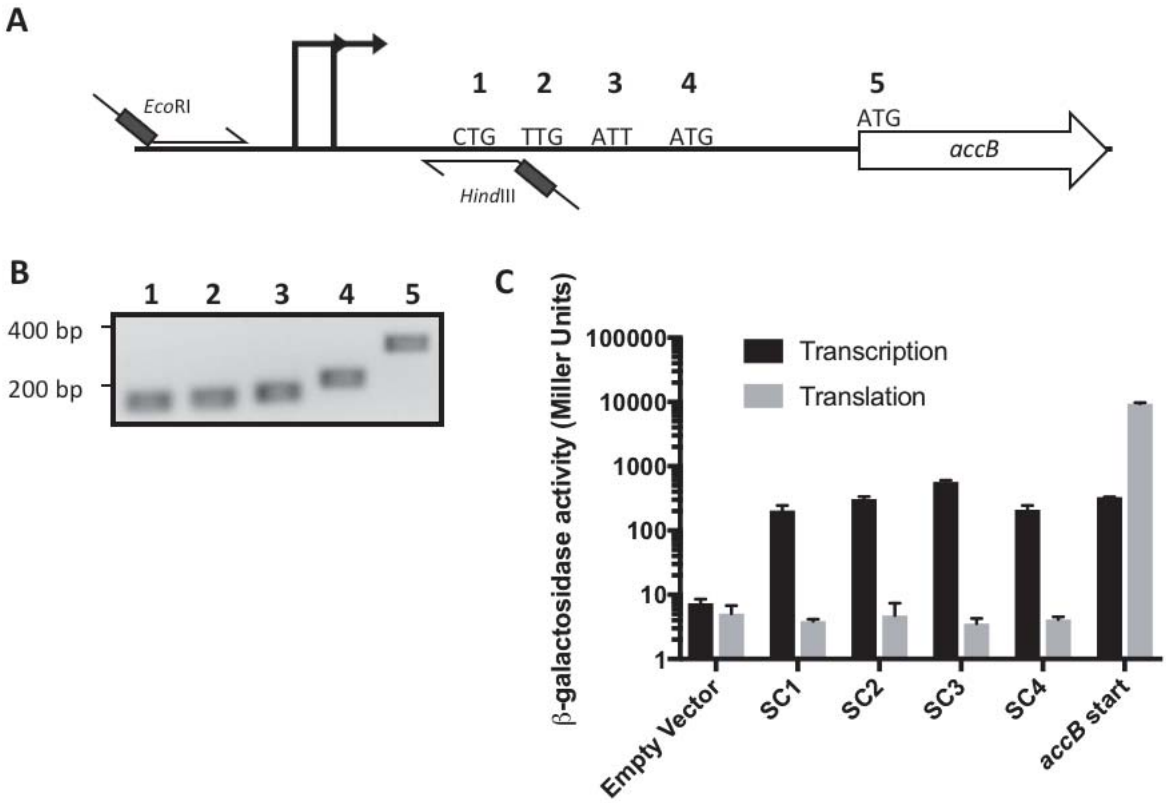

Figure 4.4 β-galactosidase activity of *accB* promoter-start codon fragments

(A) There are four candidate start codons within the largest ORF upstream of *accB* (1-4, not to scale). The primer binding sites for promoter-fragment cloning are shown for reference. (B) PCR confirmation of the PCR-amplified promoter-start codon fragments before cloning into vectors pRW224 and pRW225. (C) The β-galactosidase activity measured from all 5 transcription vectors was statistically significant compared to the empty vector (black; $P \leq 0.05$, 2way ANOVA). Only the *accB* start codon shows significant translation initiation activity (grey; $P \leq 0.05$, 2way ANOVA). Both 2way ANOVAs use the two-stage step-up method of Benjamini, Krieger and Yekutieli test for multiple comparisons. The mean and standard deviation of 3 biological replicates is shown.

deviation of biological replicates. The data show that all five promoter fragments initiated transcription (Fig. 4.4C, black), but only the *accB* start codon could initiate translation (Fig. 4.4C, grey). Under standard LB growth conditions translation was not initiated from any of the candidate alternative start codons. It is therefore unlikely that this ORF contains a CDS. Note, however, that there are smaller ORFs in the two alternative reading frames. It is therefore possible that an alternative ORF codes for an essential protein or leader peptide. Peptides as small as 2 amino acids have previously been reported in the literature as being encoded within the 5′ UTR of an operon (Islam *et al.* 2012).

4.2.4 Predicted folding of the insertion free region within the 5′ UTR

Constructs of the *accB* promoters cloned upstream of a chloramphenicol acetyltransferase reporter gene showed no changes in acetyltransferase activity whether the portion of 'bent' DNA or predicted ORF (or both, which coincide with the IFR) were deleted, nor the introduction of a stop codon into the predicted ORF (Li and Cronan 1993). This suggests that either modification to the IFR does not impact *in vitro* transcript expression and therefore the IFR may not be involved in transcript regulation, or the conditions tested were inappropriate to detect a change in expression activity.

One hypothesis is that the 5′ UTR of mRNA may regulate mRNA translation. The *accBC* operon codes for biosynthetic enzymes and a hypothesis for the function of the *accB* 5′ UTR is a riboswitch. Riboswitch ligands are typically small metabolites, however, an increasing breadth of ligands and complex riboswitch mechanisms are being discovered (Mehdizadeh Aghdam *et al.* 2016). It is possible that the 5′ UTR

functions as a sensor domain and responds to a chemical signal that mediates translation of AccB. In support of this hypothesis, the secondary structure of the 126 nucleotide sequence of the IFR was predicted using mfold (default settings; Zuker 2003) and was predicted to have a high degree of hairpin folding (Fig. 4.5). Moreover, there are two predicted stable structures that show alternate hairpin folding (green and pink; Fig. 4.5 A and C). It is conceivable that these two alternate structures might transition via an intermediate structure (Fig. 4.5B) and have a regulatory role *in vivo*. Perhaps insertion of a transposon into this region disrupts correct secondary structure folding, resulting in a constitutively 'off' operon and no expression of *accBC* which would be lethal, or a constitutively 'on' operon resulting in overexpression of *accBC* which could be toxic.

However, this hypothesis is conflicting with the reported data presented above that deletion of this region does not alter acetyltransferase activity, suggesting the IFR does not modulate downstream translation. Similarly, a 100-fold increase in gene copy number results in only a mild (2-4-fold) increase in mRNA and protein levels (S. Li & Cronan 1993) suggesting *accBC* expression is tightly controlled at the transcription level (S. J. Li & Cronan 1993).

Therefore, an alternate explanation is that the hairpins form a regulated transcription termination sequence. Within the IFR is a string of 11 nucleotides, 10 of which are thymine bases, which, when transcribed, will form a poly-U tract and might have a role in rho-independent termination. Disruption of this secondary structure by an inserted transposon might result in a constitutively 'off' *accBC* operon which would be lethal.

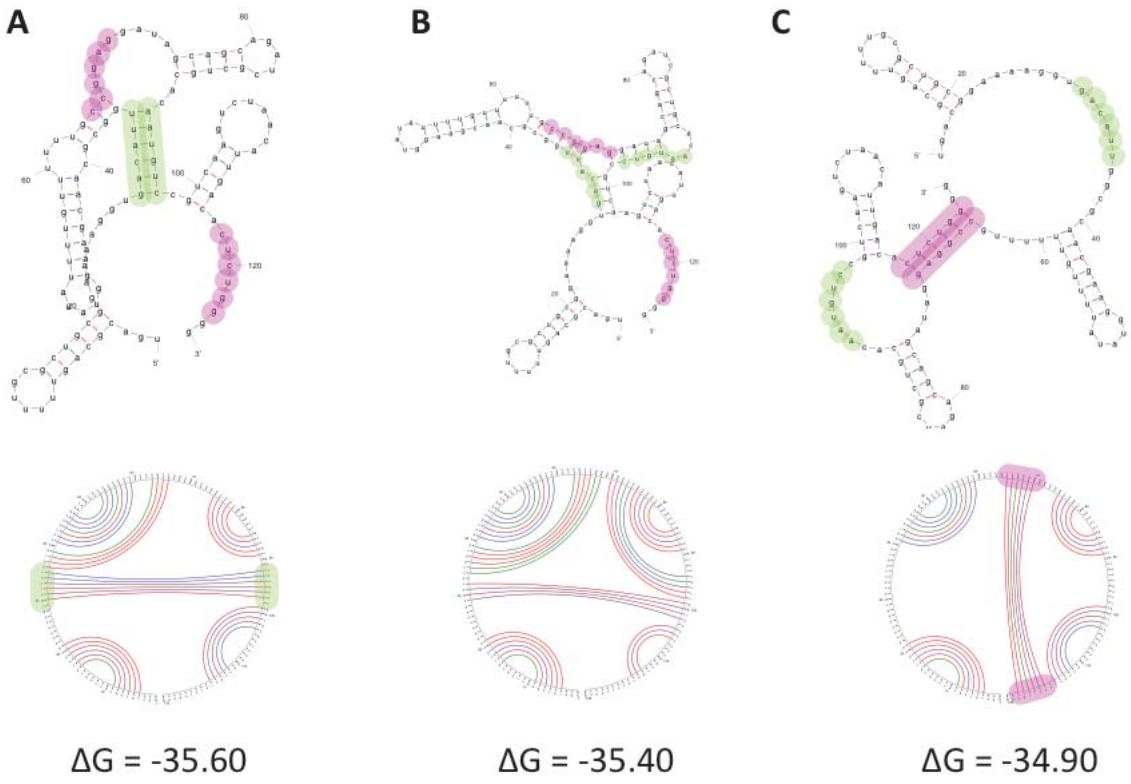

Figure 4.5 Predicted folding of the IFR within the *accB* 5′ UTR

The secondary structure of the 126 nucleotide IFR within the *accB* 5′ UTR was predicted using mfold (default parameters). (A-C) The top three results are presented here, ΔG units are kcal/mol. Two variable hairpin structures are highlighted in green (A) and pink (C).

4.2.5 Identification of transposon insertion bias at the 3′ end of transcripts with essential and non-essential regions

Another distinctive insertion profile identified within the genome is the insertion bias of transposons at the 3′ end of some essential genes, for example within *metG*, which has both an essential domain and a downstream dispensable domain (Fig. 4.6A). However, insertions within the dispensable 3′ domain are restricted to one orientation. The absence of transposon insertions in the opposite orientation to chromosomal gene expression suggests that transcription read-through from the transposon into *metG* is lethal (Fig. 4.6A). This insertion profile is not a feature of all essential genes with a downstream non-essential domain (Fig. 4.6B). Instead, this relatively rare insertion profile might suggest that disruption of a specific regulatory element is lethal. In theory, this insertion profile can be extrapolated to apply within operons of essential and non-essential genes, both mid-operon or at the 3′ end. Within the BW25113 genome this profile has only been observed in 2 operons: *rpsF* and *rpmB*; and in both examples the non-essential gene is the last gene of the transcription unit.

4.2.6 Analysis of the insertion profile within the rpsF operon

Within the TraDIS data, the *rpsF* operon is of particular interest because it is the largest transcription unit with lethal antisense transcription read-through from the transposon. In support of the hypothesis that disruption of a specific regulatory element is lethal, for both the *rpsF* and *metG* examples, antisense promoters are reported in a dRNA-seq dataset (Fig. 4.7). The absence of opposing transposon insertions might therefore reveal previously unreported tightly-regulated genetic elements.

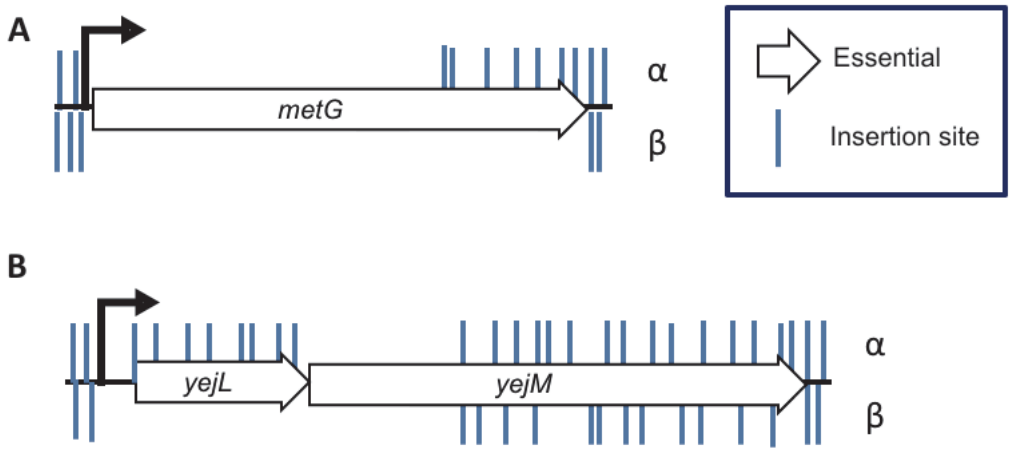

Figure 4.6 Schematic of different insertion profiles within essential genes with downstream dispensable domains

Transposon insertion sites are represented by blue bars above and below the gene track, corresponding with each orientation of the transposon. (A) Insertions within the essential gene *metG* are at the 3′ end of the CDS and restricted to the α-orientation only. Insertions in the β-orientation frame the annotated gene. (B) In contrast, insertions within the non-essential region of essential gene *yejM* are found in both orientations for the full length of the non-essential domain.

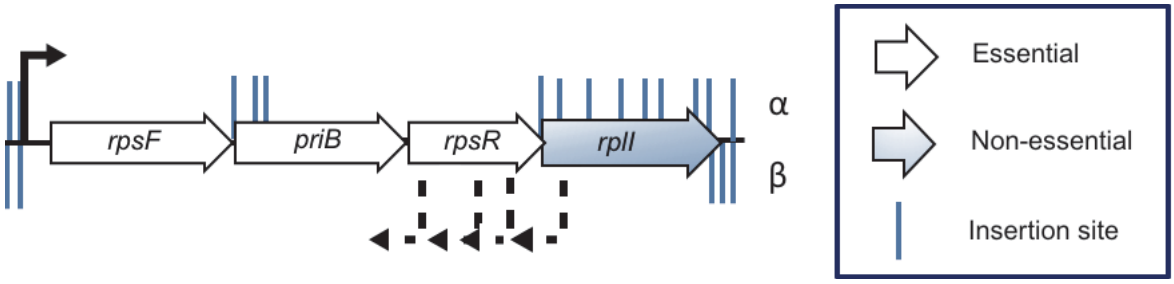

Figure 4.7 Insertion bias at the 3′ end of the *rpsF* operon transcript

Transposon insertion sites are represented by blue bars above and below the gene track, corresponding with each orientation of the transposon. Essential and non-essential genes are shown in white and blue respectively. Annotated promoters are shown with a bent arrow, predicted promoters from a dRNA-seq[3] dataset are indicated by dashed bent arrows. Transposon insertions within the non-essential gene *rplI* are restricted to the α-orientation.

The *rpsF* operon consists of three essential genes (*rpsF*, *priB* and *rpsR*) and one non-essential gene (*rplI*). Transposon insertions within *rplI* are restricted to one orientation (Fig. 4.8A). The gene *rplI* can evidently be disrupted by transposon insertions, and so the absence of transposon insertions within *rplI* in one orientation would suggest that antisense transcription read-through into the *rpsF* operon is lethal. A dRNA-Seq dataset that maps transcription start sites (TSS) reports 4 antisense TSSs within the *rpsF* operon (Fig. 4.8B; Thomason *et al*. 2015). The position of the 4 TSSs are downstream of the non-permissive transposon-insertion sites within *rplI*. One hypothesis for the absence of β-orientation transposon insertion in *rplI* is that insertion of a transposon here deregulates the activity of the native antisense promoters (Fig. 4.8B). Furthermore, there is a large ORF with a putative CDS downstream from the first predicted promoter (p1, Fig. 4.8C). These data raise two questions: is there an antisense CDS opposing the *rpsF* operon; and can TraDIS identify native antisense regulatory elements? To address this the antisense sequence of the *rpsR* and *rplI* genes were inspected for additional genomic features. Two large CDSs of 540 bp and 333 bp were identified opposite the *rplI* and *rpsR* genes, respectively. The 333 bp CDS antisense to the *rpsR* gene and downstream from the first predicted transcription start sites was of particular interest. This CDS contains two candidate AUG start codons, with two moderately AG rich regions or putative ribosomal binding sequences (RBS) upstream, and two overlapping putative -35 and -10 promoter consensus sequences (Fig. 4.9A).

To determine whether there is a nested protein coding sequence that is expressed in the opposite orientation to the *rpsF* operon, candidate promoter fragments with putative start codons were cloned into transcription and translation

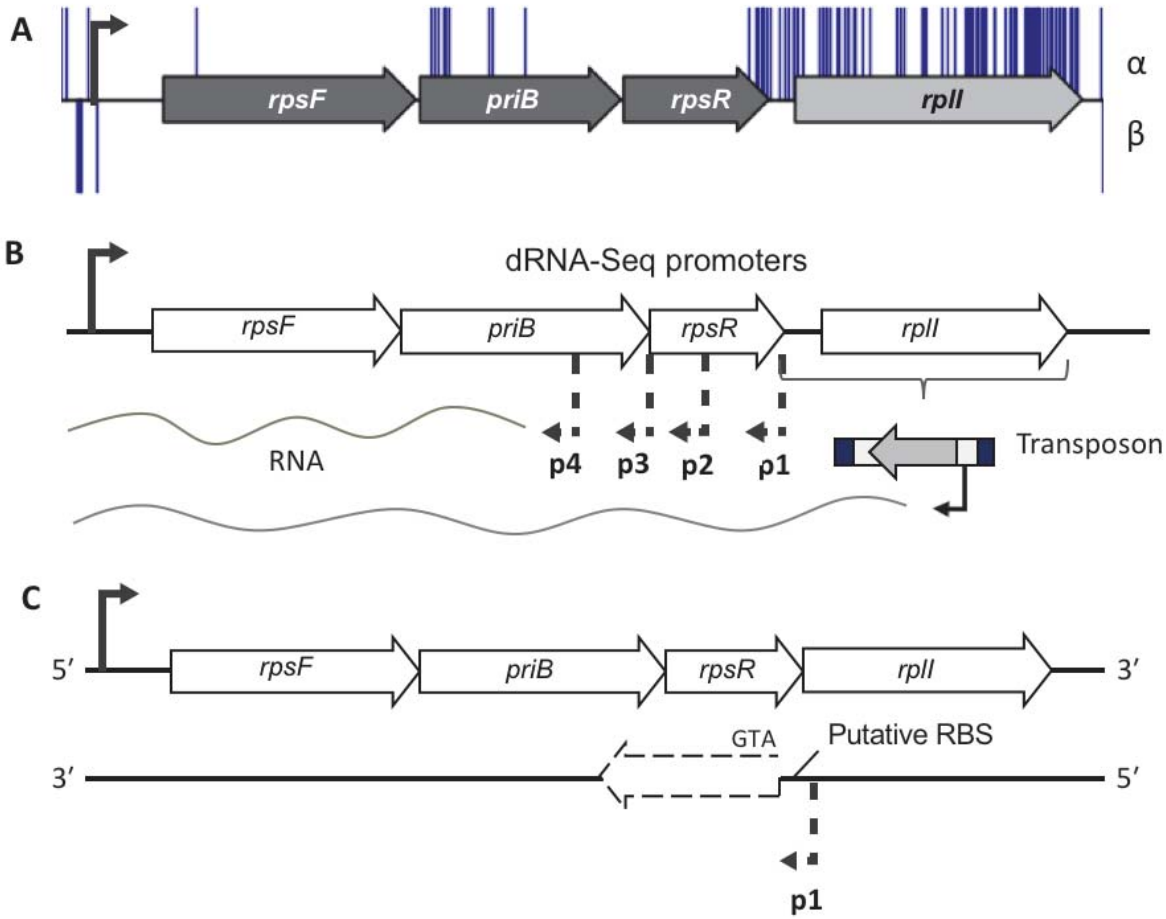

Figure 4.8 Characterisation of the *rpsF* operon.

(A) The *rpsF*, *priB* and *rpsR* genes are essential and have few transposon insertions (dark grey). The *rplI* gene is non-essential (grey), with many transposon insertions but only in one orientation. There are no transposon insertions in the orientation that would produce an antisense RNA transcript (β-orientation). (B) Putative promoters identified by the Thomason dRNA-SEQ dataset are oriented in the opposing direction to the *rpsF* operon and located within the *rpsR* and *rplI* genes (dashed lines, p1-4). Insertion of a transposon into the *rplI gene* with transcription read-out antisense to the *rpsF* operon, results in a non-viable mutant, possibly as a result of deregulation of native antisense promoters. (C) Downstream from the p1 TSS is a large ORF with a start codon and putative RBS, which could potentially be a nested CDS (dashed lines).

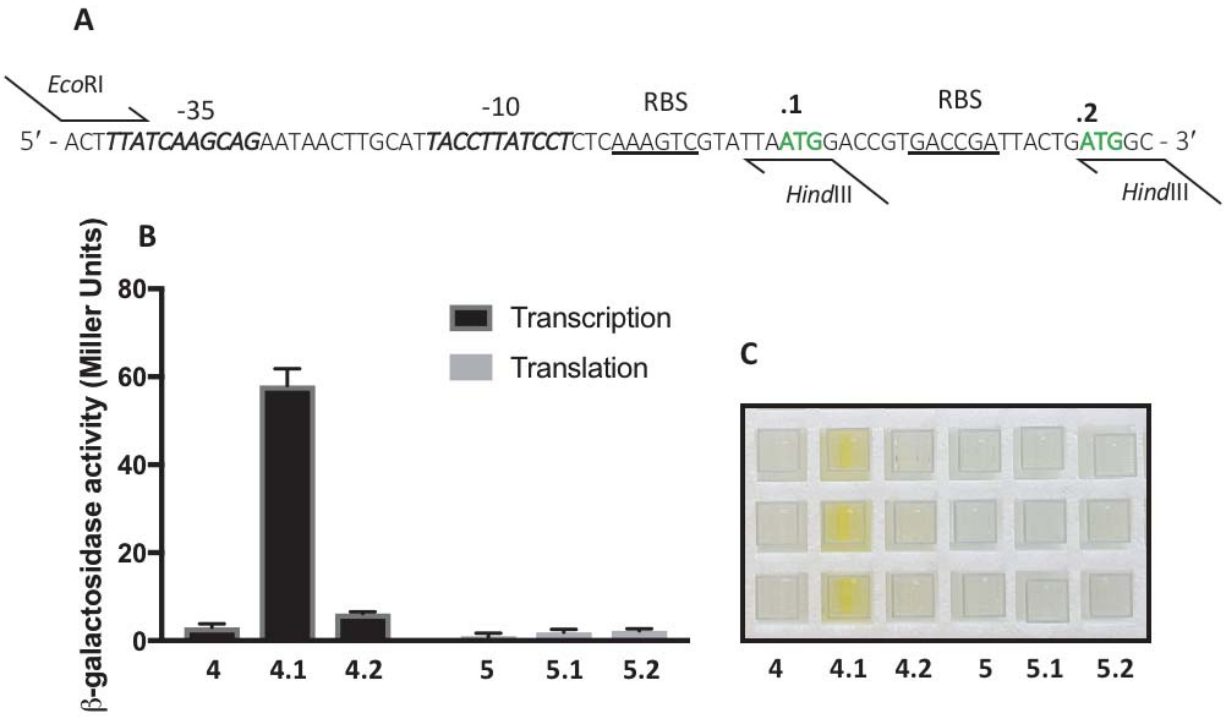

Figure 4.9 Confirmation of an antisense promoter in the *rpsF* operon.

(A) A large ORF opposing the *rpsR* gene is downstream from the first predicted transcription start site (p1) and contains two candidate start codons and putative RBSs. The sequence of this region is shown with highlighted features. Candidate start codons are shown in green (.1 and .2), putative RBS sites are underlined upstream. Two overlapping -35 and two overlapping -10 consensus sequences are shown in italics. Primers are included for reference; specific primer binding sites are not shown. (B and C) Empty vectors pRW224 and pRW225 (4 and 5 respectively) show no β-galactosidase activity. The promoter fragment '.1' but not '.2' initiated low-level transcription, but not translation (P = 0.0001, 2way ANOVA with Dunnett's multiple comparisons test). The mean and standard deviation of 3 biological replicates is shown.

expression vectors pRW224 and pRW225, respectively, upstream of *lacZ*. The promoter fragments included ~70 nt upstream of the putative -35 element, and 2 codons after the start codon of interest (indicated by the *Hin*dIII primer binding site, Fig. 4.9A). The reporter vectors were transformed into BW25113 and transformants grown in LB at 37°C. Samples were collected during the exponential phase of growth and β-galactosidase activity was measured to determine the transcriptional and translational capacity of the candidate promoter fragments and start codons. No β-galactosidase activity was detected for either the empty vector, or the translation-initiation vectors. It is possible that the putative RBS motifs are not functional and are merely annotation artefacts, in the absence of an appropriate positive control a conclusion cannot be established from these data. However, weak β-galactosidase activity for one of the promoter fragments (p4.1) was detected, confirming the presence *in vitro* of antisense promoter activity (Fig. 4.9B and C). This was surprising, as the p4.1 and p4.2 promoter fragments are identical, except for an additional 21 nt at the 3′ end of the p4.2 fragment.

These data confirm the presence of functional promoter sequences opposing *rplI* transcription, in support of data presented by Thomason *et al*. (2015), and suggest that these promoters are either tightly regulated, or the p4.2 transcript is highly unstable. The need for tight regulation might support the theory that deregulated antisense transcription is lethal within this operon. Nonetheless, there are two alternative explanations for the presence of antisense transcription: either the transcription is spurious, usually as a result of AT rich DNA (Singh *et al*. 2014); or there is a functional antisense RNA.

4.2.7 Quantification of the AT content of the rpsF operon

Antisense transcription within a gene has been cited in the literature as a pervasive artefact of AT rich DNA (Singh *et al.* 2014). Therefore, the AT/GC content of the *rpsF* operon was examined to determine whether the AT content of the genes *rpsR* and *rplI* might explain the presence of native antisense transcription. The total GC content of the *E. coli* K-12 genome is 50.8% (Hayashi *et al.* 2006). Within the *rpsF* operon, for a fixed window size of 100 nt, the percentage GC fluctuates mildly but does not deviate hugely from the genome mean (Fig. 4.10). This suggests that the antisense promoters are not an artefact of AT rich DNA, however any functional relevance *in vivo* has yet to be shown.

A hypothesis for the absence of β-orientation transposon insertions within the *rplI* gene is that production of a complementary antisense sRNA silences the sense transcript and is therefore lethal. However, this raised the questions: why is this insertion profile not observed for all essential genes; what is unique about this operon? The explanation might simply be due to comparative promoter strengths of the native promoters of essential genes and the opposing promoter strength of the kanamycin resistance cassette. Essential genes might have stronger promoters than the transposon promoter and so the effect of a silencing antisense transcript is negligible. Alternatively, reports state that the effect of clashing promoters is greatest over longer distances, especially when the difference between promoter strength is marginal (Georg & Hess 2011; Sneppen *et al.* 2005); this could be the case with the *rpsF* operon. But the verification of at least one native internal antisense promoter raises yet more questions: what is the purpose, if any, of these promoters and why aren't they lethal *in vivo*?

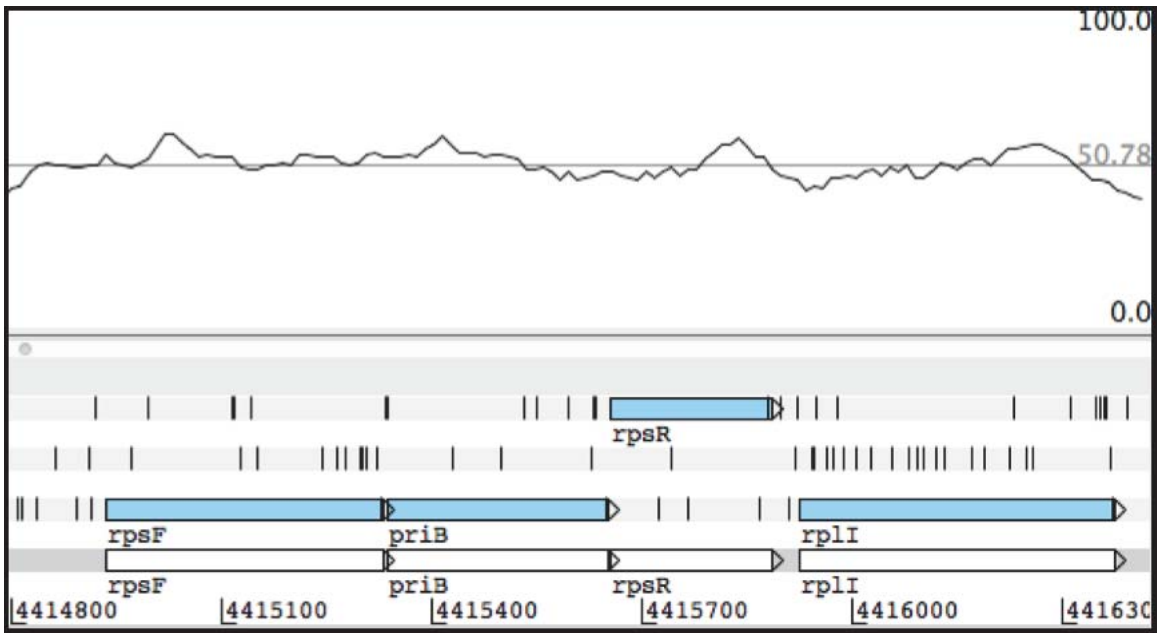

Figure 4.10 Percentage GC of the *rpsF* operon

The total GC content within a given sliding window of 100 nt is shown for the *rpsF* operon. The mean GC content (50.78%) of the genome is shown as a fixed horizontal line in grey. The GC content of the *rpsF* operon does not fluctuate from the genome mean.

Functional antisense transcription was first documented in bacteria with the discovery of the ColE1 plasmid replication initiation mechanism in 1980 (Itoh & Tomizawa 1980). Initially, examples like this were thought to be rare exceptions. However, with the development of NGS-based technologies to map transcription start sites the widespread prevalence of internal promoters was revealed; internal promoters were identified both in frame with and antisense to existing annotated mRNA transcription (Thomason *et al*. 2015b). Some reports suggest that these transcripts are spurious as a result of AT rich DNA (Singh *et al*. 2014; Raghavan *et al*. 2012), while many antisense RNAs have been discovered to have diverse functional roles (Opdyke *et al*. 2004; Kawano *et al*. 2007; Fozo *et al*. 2008). In many cases the function of an antisense RNA is to regulate the expression of overlapping genes (Dornenburg *et al*. 2010). One possible theory to explain the *rpsF* operon TraDIS data is that the native antisense promoters regulate expression of the sense essential genes, a mechanism which is deregulated by insertion of a transposon upstream. These findings may have inadvertently identified a new regulatory mechanism for the expression of specific ribosomal genes, however, this remains to be verified.

4.2.8 Discovery of statistically significant transposon-free regions.

As shown in Chapter 3, the frequency of transposon insertions was very low, or zero, in the recently annotated genes *ynbG*, *yobI* and *yqcG* in the BW25113 reference genome, which code for small proteins YnbG, YobI and YqcG (Fig. 3.9C, D and E). Small proteins are classically overlooked when annotating a genome. Reasons for this include that it is difficult to detect sequence conservation or homology for a given short sequence or because they are hard to validate biologically because they are

unstable, they are expressed transiently, or are conditionally expressed. In addition, the NCBI requires a minimum of 200 nt for a sequenced feature to be deposited in the database. Therefore, it has been hypothesised that genes encoding small proteins are underrepresented in curated genomes and more almost certainly remain to be identified. It is only possible to detect new genes among a TraDIS dataset if they are essential (or conditionally-essential) as there is a measurable change in insertion frequency, limiting the number of new genes this method can identify. Nonetheless, the insertion data were inspected for additional transposon-free intergenic regions that might correspond with new protein CDSs. Given that for this dataset the threshold of a statistically significant IFR within a gene is 47 bp (ρ_{gene} of 0.005), a window size of 47-200 nt was selected. Using a custom script to interrogate the insertion data for IFRs ranging from 47-200 nt, 3404 IFRs were identified throughout the genome. Using the program IntersecBed, 666 of these IFRs were exclusively intergenic and do not overlap with any annotated CDS of the annotated BW25113 genome (CP009273.1; Quinlan and Hall 2010). This approach is conservative and does not include any intergenic IFRs that overlap with the 5′ or 3′ end of a CDS. The total data were then inspected by eye to identify IFRs that might coincide with ORFs. An initial shortlist of 30 ORFs was compiled for further investigation. The ORFs were inspected for potential start codons and cross-referenced against a transcription start-site dataset and the Tjaden intergenic transcriptome data (Tjaden *et al.* 2002; Thomason *et al.* 2015). This resulted in a shortlist of ten examples of putative CDSs selected for further characterisation (Fig. 4.11; Table 4.1). While these putative genes were not investigated further, they are promising candidates for novel genes. Canonical start codons were found in 8 of these sequences, 5 of the candidates have been previously

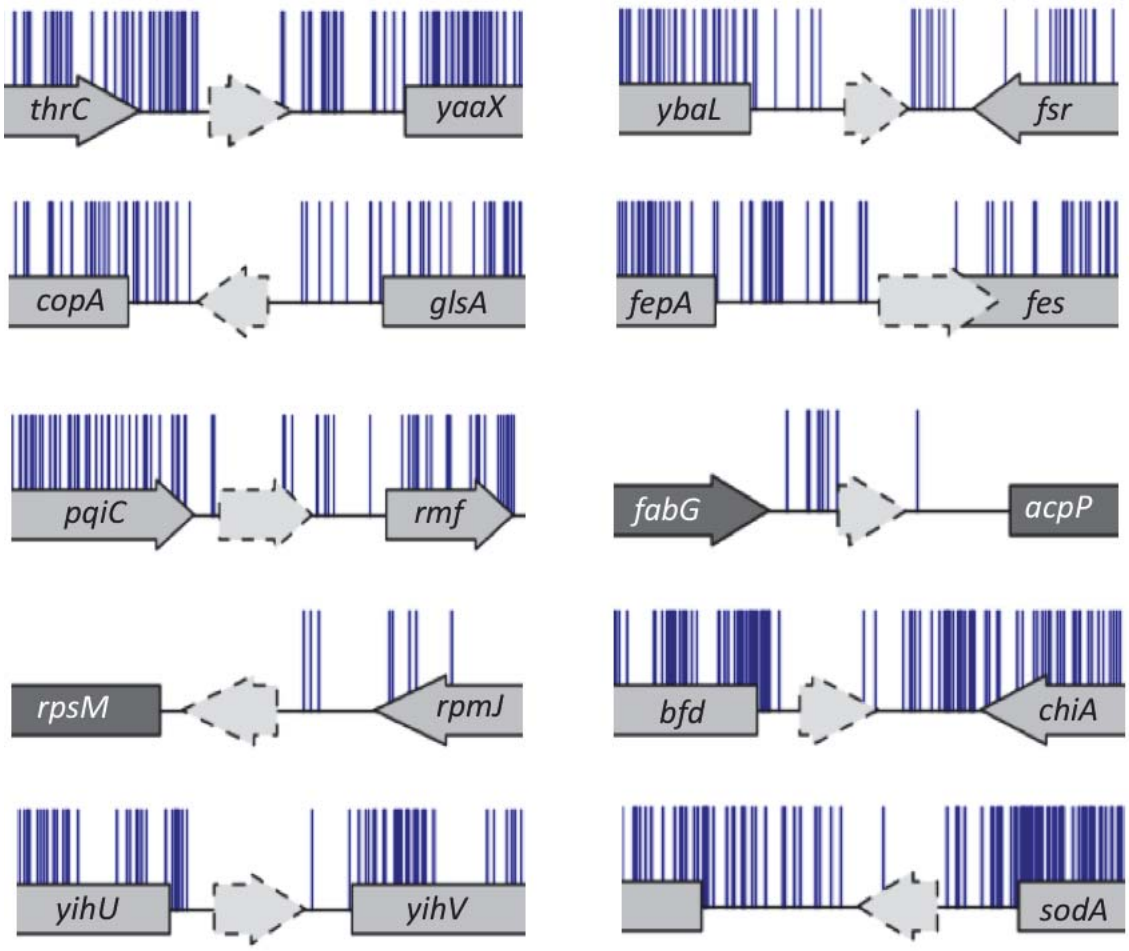

Figure 4.11 Putative essential short coding sequences

The insertion profiles of 10 putative genes. Individual figures are to scale, but not proportional to each other. Total transposon insertions (both orientations) are indicated by dark blue lines, non-essential genes shown in light grey and essential genes shown in dark grey. Putative CDSs are indicated by dashed outlines. Each of the candidate new genes coincides with a significant IFR.

Table 4.1 Candidate short ORFs identified from the insertion data

Up. gene	Down. gene	Start codon	Size (bp)	Tjaden data	Sequence
thrC	*yaaX*	AUG	84		MQPGFFMKKLWRKMTGKKEKFSINAVT
ybaL	*fsr*	AUU	72		IIEFCRKMELAAKKARIGKQKAV
copA	*glsA*	AUG	78	5′ UTR	MSNPDANRADFLIFTDFYPPGLLIS
fepA	*fes*	AUG	105	sRNA/ORF	MQIVINNIINIFLQSMKNCTVNMGLWCD -GVKSRK
pqiC	*rmf*	AUG	126	Unknown function	MLPEAVFLSLTCGKICSSSHFLYNRHARV-AHKYDSGVNFAH
fabG	*acpP*	GUG	60	sRNA/ORF Unknown function	VQNDLRYWGVRPQNNVKSW
rpmJ	*rpsM*	AUG	69	Unknown function	MSVRFHLSILKTGFSAWNVHIK
chiA	*bfd*	AUG	63		MKIVLISITERLFFKQRIEI
yihU	*yihV*	AUG	87		MTCFACLFLIVFVIYRQKIDSRHFLNNW
rhaT	*sodA*	AUG	75		MILKMIINAVLFVRVWFCRKMPEM

reported by Tjaden *et al.* (2002) in a transcript hybridisation microarray assay, and 5 are downstream from predicted transcription start sites identified by Thomason *et al* (2015).

4.3 Conclusion

The above examples show that in addition to categorizing conditionally essential genes, detailed analysis of a high-throughput mutagenesis study can reveal information about genome annotation and transcription regulation. The TraDIS data revealed a significantly large essential region within the unusually large UTR of the *accBC* operon. Additional inspection of the TraDIS data revealed a unique insertion profile opposing certain essential genes, and characterisation of one of these examples resulted in the *in vitro* identification of an active promoter fragment, which, if functional *in vivo* potentially reveals a new mechanism of ribosomal gene expression regulation. Lastly, 666 IFRs were identified exclusively between annotated genes. The intragenic regions of 6 bacteria, including *E. coli*, have recently been reported to be selectively conserved, implying important functional roles of these regions (Thorpe *et al.* 2016). Coupled with the TraDIS data presented in this chapter, these suggest that there are further essential features of the *E. coli* genome to be characterised.

The main objective within this chapter was to search for new essential genes within the reported IFRs, but there are alternate explanations that might account for an insertion free region such as an essential sRNA, a DNA-bound protein that excludes transposon insertion, or an essential promoter (Christen *et al.* 2014; Barquist *et al.* 2013; Manna *et al.* 2007). In addition, there are specific known motifs within the genome that cannot be disrupted by transposon insertions in our dataset. For

example, the principle 4 ter sites: terA, terB, terC and terD; are located within intragenic regions, none of which are disrupted by transposon insertion, and 3 are within statistically significant IFRs. Ter sites are directional motifs that, when bound by the protein Tus, trap the DNA replication fork ahead of the ter site to terminate replication (Khatri *et al.* 1989). Targeted point mutations within these ter sites lower or abolish replication arrest in *in vitro* studies of plasmid replication (Coskun-Ari & Hill 1997). It is possible that within the identified statistically significant IFRs, there are novel motifs that have an essential role in *E. coli*.

Finally, the lack of identified insertions within specific regions may simply be an artefact of the protocol. Identification of transposon junctions by sequencing requires amplification of the intended target by PCR. It is possible that certain genomic regions might be disrupted by a transposon but, due to a highly folded secondary structure, fail to be amplified by PCR.

Data presented in this chapter highlight the power of a highly saturated transposon library. Transposon screens have the capacity to not only identify essential genes but also the complex regulatory elements that surround them. The simplified operon model of bacterial gene expression and regulation is no longer sufficient to capture the complexities of bacterial genetic regulation, as a breadth of regulatory mechanisms are discerned. Perhaps TraDIS could be adapted for the discovery of specific novel genetic elements, if the transposon was designed in such a way to discern them.

Chapter 5

Identification of Conditionally Essential Genes Required for Growth in Polymyxin B

5.1 Introduction

The transposon library presented in the previous chapters is the most dense library published. As such, this library is a powerful tool for high-throughput screening of antibiotics to identify the regions of the genome that, when disrupted, render a mutant more, or less, susceptible to an antibiotic of interest. The outcome of a TraDIS experiment is dependent upon the experimental design. A 'high' antibiotic concentration will result in enrichment of mutants that are resistant to the antibiotic relative to the parent. In contrast, use of a 'low' antibiotic concentration will identify genes required for survival in the presence of the antibiotic, termed 'conditionally essential' genes. This chapter reports the use of our library to identify mutants that are more susceptible to polymyxin B.

The antibiotic polymyxin B was chosen for 2 reasons. First, it is often described as a 'drug of last resort' and therefore of clinical importance (Zavascki *et al.* 2007). Secondly, polymyxin B targets the cell envelope, which is often described as the 'first line of defence' to antimicrobial compounds and an inherently intrinsic form of resistance (Mares *et al.* 2009). Understanding the network of proteins that contribute to the construction and maintenance of this structure is fundamental to understanding how to target the cell envelope.

Polymyxin B, originally isolated from a garden in Surrey and named 'aerosporin', is synthesised by *Bacillus polymyxa* and comprises a mixture of structurally related compounds (Ainsworth *et al.* 1947; Paulus & Gray 1964). Polymyxins B1 and B2 make up the majority of polymyxin B. The structure of polymyxin B1 is shown in figure 5.1A. The overall polymyxin B structure consists of a closed ring with 5 positively charged diaminobutyric acid (Dab) residues and a fatty acyl tail, resulting in an amphipathic molecule. The positively charged Dab residues

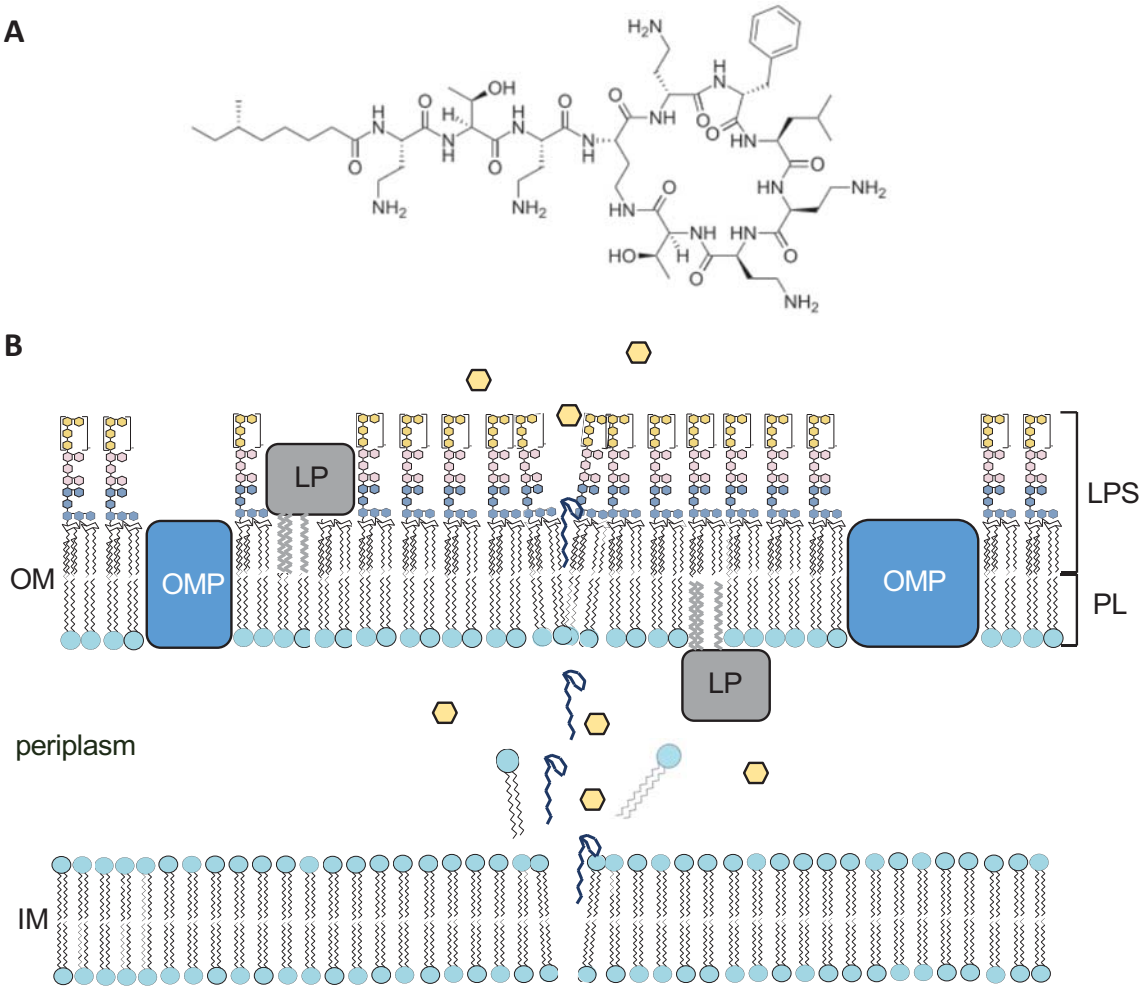

Figure 5.1 The structure and model of the polymyxin B mechanism of action

(A) Polymyxin B is an amphipathic cation; it has both positively charged, hydrophilic residues and a hydrophobic fatty acyl chain. (B) Polymyxin disrupts the cell membrane via a 2-step mechanism. The positively charged residues of polymyxin B interact with the phosphate groups of lipid A via electrostatic interaction. The hydrophobic acyl tail then inserts into the outer membrane and disrupts membrane stability. This facilitates auto-uptake of polymyxin B, where it can subsequently insert into the inner membrane. Membrane destabilisation results in the leakage of soluble cell components and cell death.

interact electrostatically with the negative phosphate groups of lipid A within LPS, displacing divalent cations that cross-link LPS molecules and disrupting the LPS packing of the OM (Peterson *et al.* 1985; Kanazawa *et al.* 2009; Srimal *et al.* 1996). After divalent cation displacement the fatty acyl chain inserts into the membrane facilitating auto-uptake of the polymyxin molecule. The molecule then inserts into the IM and disrupts phospholipid packing, resulting in leakage of cytoplasmic small molecules and ultimately resulting in cell death (Fig. 5.1B; Vaara & Vaara 1983; Teuber & Bader 1976; Berglund *et al.* 2015; Deris *et al.* 2014). While the main mode of action is generally understood, polymyxin B is also reported to have off target secondary sites of action that are poorly understood (Deris, Akter, *et al.* 2014).

Mechanisms of resistance to polymyxin B often involve enhancing the barrier protection of the OM, limiting polymyxin access to the IM and cytoplasm. Possibly because once inside the cell the effects of polymyxin B are pleiotropic and therefore multiple adaptations would be required to combat these. A common adaptive mechanism of resistance observed in clinical isolates is modification to LPS to decrease the negative charge, weakening the strength of interaction between polymyxin B and LPS (Trimble *et al.* 2016).

The aim of the work presented in this chapter was two-fold: to screen a transposon library of *E. coli* strain BW25113 to show as proof of principle that TraDIS can be used to identify conditionally essential genes required for growth under a specific condition. And secondly, to identify the conditionally essential genes required for survival in the presence of sub-inhibitory concentrations of polymyxin B. Assessment of the role of acquired resistance, or the role of core-essential genes, is outside the scope of this chapter.

5.2 Results and Discussion

5.2.1 Determination of growth conditions that partially inhibit growth in the presence of polymyxin B

Preliminary experiments were designed to determine the minimum concentration of polymyxin B (hereafter referred to as MIC) that would inhibit growth of the strain BW25113 under experimental conditions used for screening a TraDIS library of mutants. An amended MIC protocol was used for initial experiments (Andrews, 2001). Three biological replicates of BW25113 were inoculated into a 96-well plate containing LB supplemented with 2-fold dilutions of polymyxin B in the range 0.008 to 4 μg/ml. The optical density (OD_{600}) of cultures was recorded after 16 h of overnight growth at 37°C with aeration. Growth was detected for all three replicates in the presence of polymyxin B at a concentration of 0.25 μg/ml and below (Fig. 5.2A). In contrast, no growth was detected in the presence of 1 μg/ml or higher concentrations of polymyxin B. Growth in LB supplemented with 0.5 μg/ml polymyxin B varied between replicates: no growth was detected for one out of three replicates. The result from this preliminary experiment revealed only an approximate range of polymyxin B concentrations that would inhibit growth, but did not determine the MIC.

As the endpoint data were inconclusive as to whether 0.5 or 1 μg/ml is the true MIC, the protocol was extended to monitor growth throughout incubation at 30 min intervals. Concentrations of polymyxin B ≥ 2 μg/ml, inhibited growth of *E. coli* BW25113 (Fig. 5.2B). In contrast, at concentrations of polymyxin B ≤ 0.5 μg/ml, BW25113 had similar growth kinetics to growth in LB without the antibiotic.

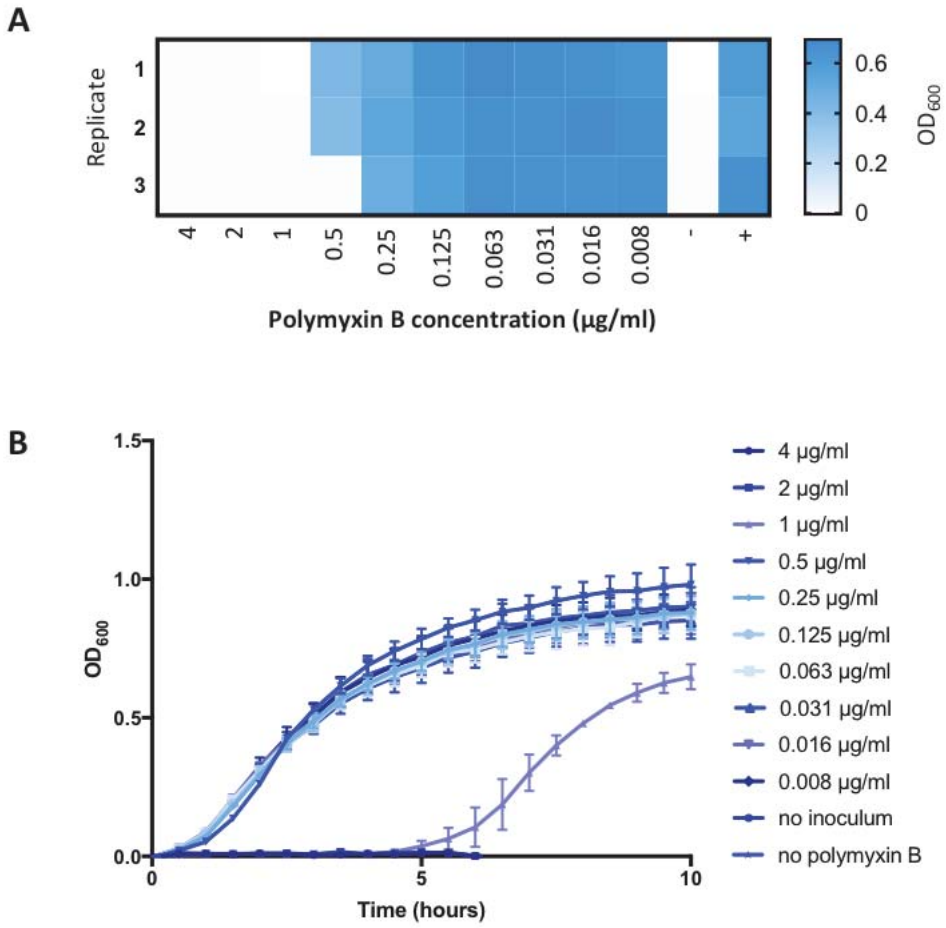

Figure 5.2 Growth of BW25113 in varying concentrations of polymyxin B

Three biological replicates of BW25113 were grown in LB supplemented with 2-fold dilutions of polymyxin B from 4 µg/ml to 0.008 µg/ml in a 96-well plate. An uninoculated negative control is denoted by '-' and a positive-control of LB without polymyxin B is denoted by '+'. (A) The endpoint optical density was recorded after 16 h. The amount of growth is indicated in blue, darker blue corresponding with a higher endpoint OD_{600} and white corresponds with no growth. (B) The OD_{600} was recorded at 30 min intervals, and the mean is plotted and s.d. shown by error bars.

The growth of BW25113 was consistent between biological replicates grown at the same time but not between experimental replicates; results varied between days whether 2, 1 or 0.5 μg/ml polymyxin B was the MIC for growth in a plastic 96-well plate, but BW25113 consistently grew in LB supplemented with polymyxin B at concentrations ≤ 0.25 μg/ml.

5.2.2 Extension of the growth conditions to reflect the parameters of a TraDIS growth experiment

The aim of the TraDIS experiment was to determine genes required for survival in the presence of polymyxin B. Therefore, to determine a suitable sub-MIC concentration of drug for TraDIS screening, BW25113 was grown from an initial OD_{600} = 0.05 in 50 ml LB supplemented with polymyxin B ranging from 0.2 μg/ml to 0.5 μg/ml, (Fig. 5.3). At increasing concentrations of polymyxin B the onset of exponential growth was delayed; this delay was most pronounced at 0.5 μg/ml. The growth curve of BW25113 in the presence of 0.2 μg/ml polymyxin B was similar to growth in the absence of any antibiotic. Therefore, as the highest concentration of antibiotic that had no measurable effect on the growth of the WT strain under TraDIS growth parameters, 0.2 μg/ml was the concentration chosen for subsequent TraDIS analysis.

To identify genes required for growth and survival in polymyxin B, two technical replicates of the transposon library were grown in 50 ml LB supplemented with 0.2 μg/ml polymyxin B. Two technical replicates of the transposon library were also grown in 50 ml LB supplemented with 0.1 μg/ml polymyxin B to identify mutants with a mild polymyxin sensitivity phenotype. The library was grown from a starting OD_{600} of 0.05 for 4 to 5 generations to a final OD_{600} = 1.00 and harvested following the standard TraDIS protocol.

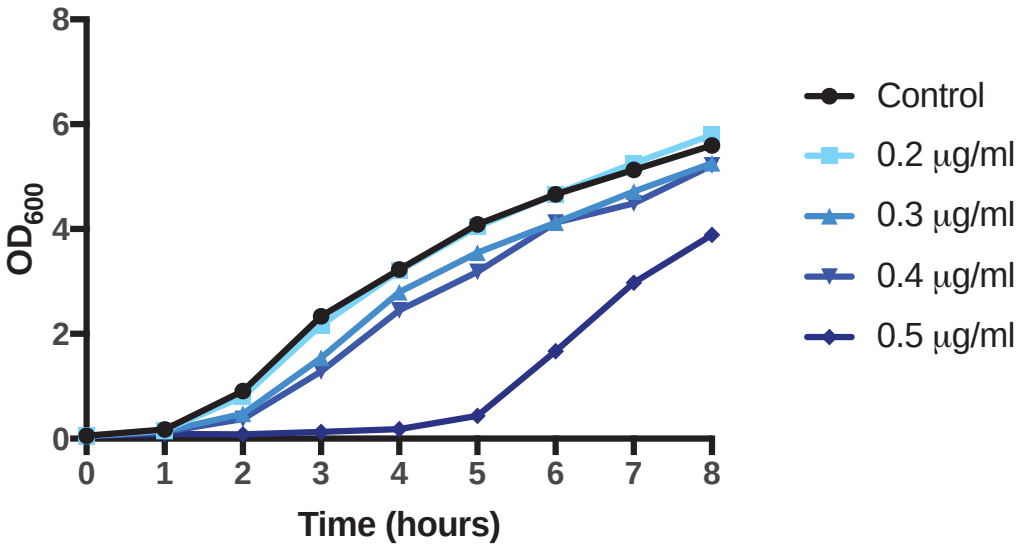

Figure 5.3 Growth of BW25113 in the presence of polymyxin B under the same conditions as those used for TraDIS experiments

BW25113 was grown with aeration at 37°C in 50 ml of LB supplemented with 0.2-0.5 μg/ml polymyxin B in 250 ml flasks (blue). Growth of BW25113 in LB without polymyxin B is shown in black. The OD_{600} was recorded at 1 h intervals for 8 h.

5.2.3 Calculation of the number of mapped sequencing reads required for comprehensive mutant sampling

Having harvested samples for sequencing, it was necessary to determine an appropriate number of sequencing reads needed for sufficient sampling of the data. With an approximate known maximum number of unique insertion points (UIP) within the transposon-library it is possible to approximate the amount of sampling required to identify a given proportion of the dataset using the equation $I = s - s(\frac{s-1}{s})^n$. The sample size 's' is 901,383 (Chapter 3), 'n' is the number of sampling iterations (which equates to sequencing reads), and 'I' is the number of new insertions identified.

Plotting this equation for 0-5 million iterations (mapped sequence reads) shows the rate of discovery of new UIPs (Fig. 5.4A). Plotting the number of identified UIPs as a proportion of the total UIPs shows the percentage of UIPs identified per million mapped reads (Fig. 5.4B). This graph can be used to determine the number of mapped reads required to sample a given percentage of the total data. These values are reported in Table 5.1. For example, to ensure sampling of 99% of the possible UIPs, ~4.151 million mapped reads are required. Therefore, samples were sequenced until a combined total of 4.15 M mapped reads (per condition) were collected.

5.2.4 Sequencing and comparison of independent replicates of the transposon library

The two technical replicates of the transposon library exposed to polymyxin B were sequenced for each condition: 0.1 or 0.2 µg/ml polymyxin B. Data were checked for the presence of an inline index barcode, followed by the transposon sequence in two steps as in Chapter 3. Data recovered are summarized in Table 5.2.

The insertion index scores of each gene were calculated for both technical replicates of each condition. The IISs were compared between technical replicates for

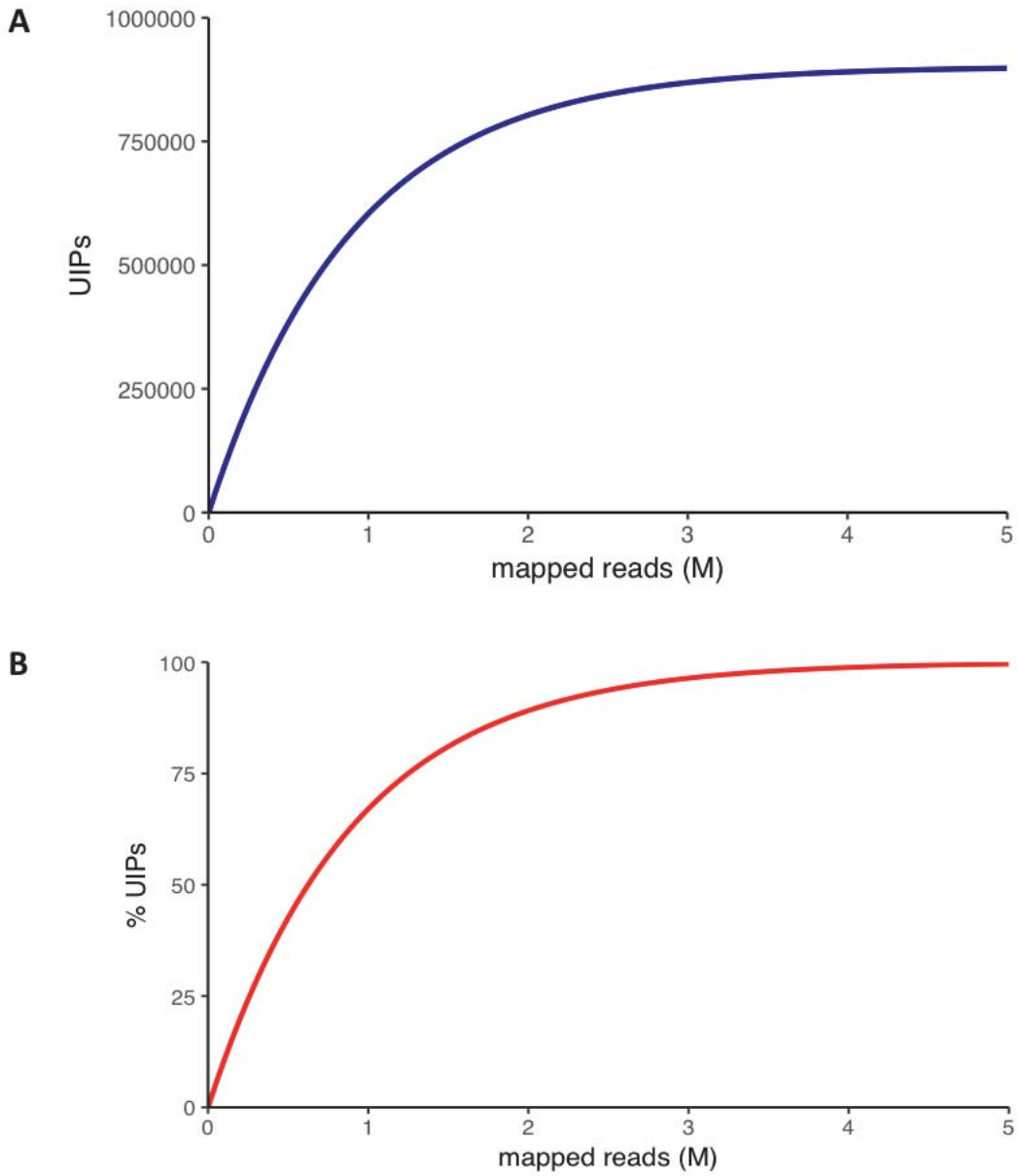

Figure 5.4 Calculation of the number of reads required to sample a given proportion of the library

(A) The equation $I = s - s(\frac{s-1}{s})^n$ was used to estimate the number of mapped reads required to sample a given proportion of the dataset, assuming a finite total number of insertions of 901,383. I = insertions, n = number of mapped reads. Plotting the equation $I = s - s(\frac{s-1}{s})^n$ from n=1 to n=5M produces the graph presented above. (B) This data was used to calculate the approximate percentage of unique insertions identified for a given number of mapped reads.

Table 5.1 The approximate number of mapped reads required to sample a given percentage of the total data

Percentage	UIPs	Number of reads
90	811245	~2.08 M
95	856314	~2.70 M
99	892369	~4.15 M

Table 5.2 The number of reads at each stage of the data processing pipeline for technical replicates of transposon library exposed to polymyxin B

Sample	Inline barcode	Tn1 check	Tn2 check	Mapped reads	UIPs
0.1 µg/ml					
Rep 1	2,894,845	2,790,801	2,541,047	2,448,085	498,977
Rep 2	2,262,290	2,206,441	2,073,605	1,981,579	545,162
Combined					**724,512**
0.2 µg/ml					
Rep 1	2,432,870	2,399,988	2,253,053	2,170,064	516,484
Rep 2	2,562,712	2,515,024	2,384,055	2,302,062	484,790
Combined					**704,413**

both polymyxin datasets as an internal measure of quality control. For both polymyxin B datasets there was a high correlation coefficient between technical replicates, R^2 = 0.98 and 0.97 for 0.1 and 0.2 µg/ml polymyxin B, respectively, and so technical replicate data were pooled (Fig. 5.5). This resulted in a total of 4,429,664 mapped reads and 724,512 UIPs, and 4,472,124 mapped reads and 704,413 UIPs for the 0.1 and 0.2 µg/ml polymyxin B datasets, respectively.

5.2.6 Identification of essential and conditionally-essential genes required for growth in the presence of sub-inhibitory concentrations of polymyxin B

The frequency distribution of insertion index scores was plotted and followed a bimodal distribution (as in Chapter 3; Fig. 5.6). These data were used to predict gene essentiality by determining the likelihood of each gene belonging to each modelled mode (S. Table 5.1). The same statistical criteria were used for this and subsequent analyses as those used in the previous TraDIS experiments.

In contrast to the 356 genes that were reported as essential in the LB-only outgrowth dataset described in Chapter 3, this increased to 398 for cultures grown in the presence of 0.1 µg/ml of polymyxin B. In the presence of a higher concentration of polymyxin B, this number increased further to 419. A three-way comparison between the identified essential genes revealed a core set of 348 essential genes, and a further 75 genes conditionally essential genes thought to be important for survival in the presence of polymyxin B (Fig. 5.7, S. Table 5.2). Of the 75 candidate conditional essential genes, 7 genes are unique to the 0.1 µg/ml polymyxin dataset and on closer inspection are all false positives (all 7 genes have the same insertion profile under all three conditions). The genes identified only in the LB outgrowth dataset, or LB

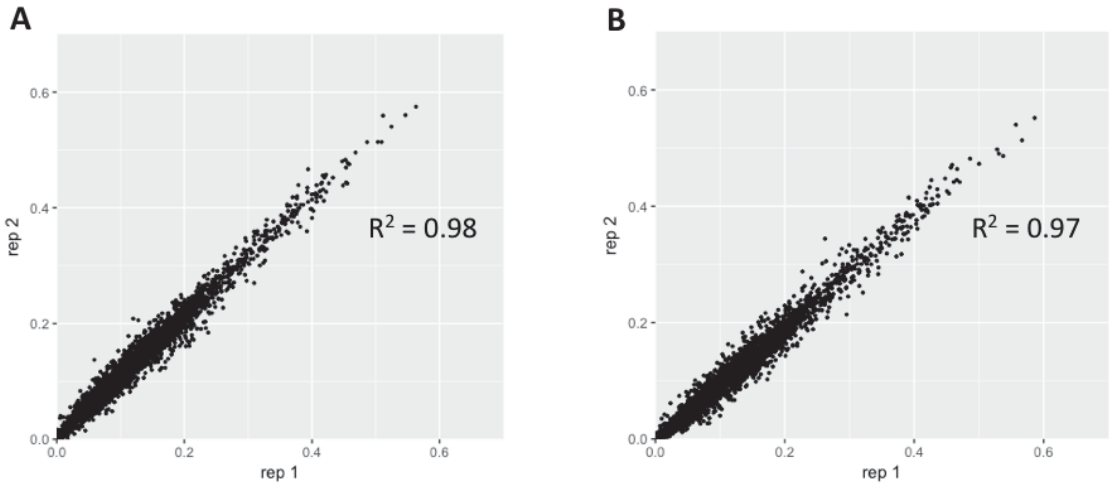

Figure 5.5 Comparison of gene insertion index scores between technical replicates

The correlation coefficient of gene insertion index scores of two independently treated samples of transposon library exposed to 0.1 μg/ml polymyxin B (A) and 0.2 μg/ml polymyxin B (B).

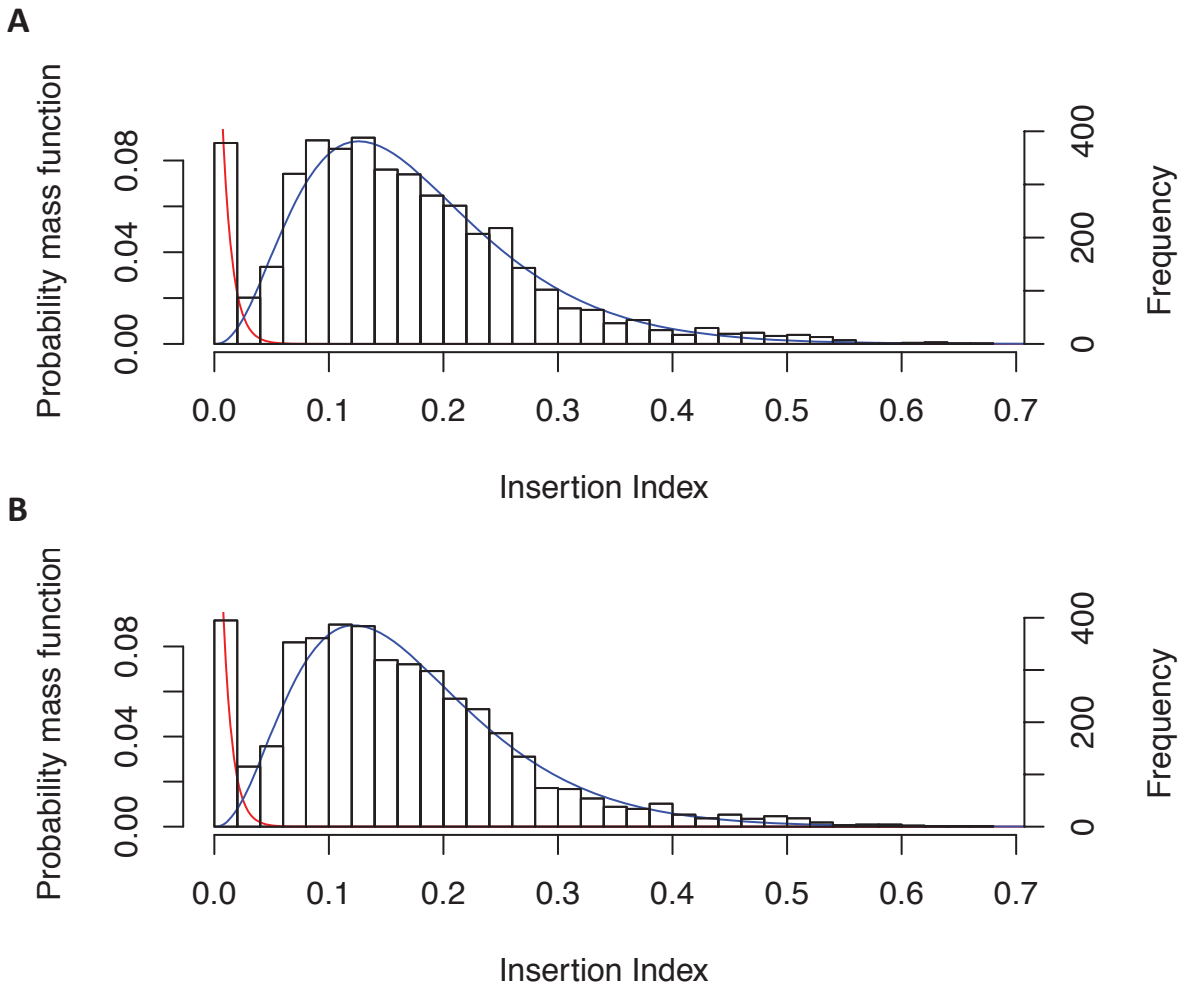

Figure 5.6 Frequency of gene insertion index scores in a transposon library following exposure to 2 concentrations of polymyxin B

The frequency distribution of gene insertion index scores for the transposon library grown in LB supplemented with 0.1 µg/ml (A) and 0.2 µg/ml (B) polymyxin B respectively. The Insertion index scores for both data sets followed a bimodal distribution. An exponential distribution model was fitted to the left mode that includes essential and conditional essential genes (red) and a gamma distribution model was fitted to the right, non-essential mode (blue).

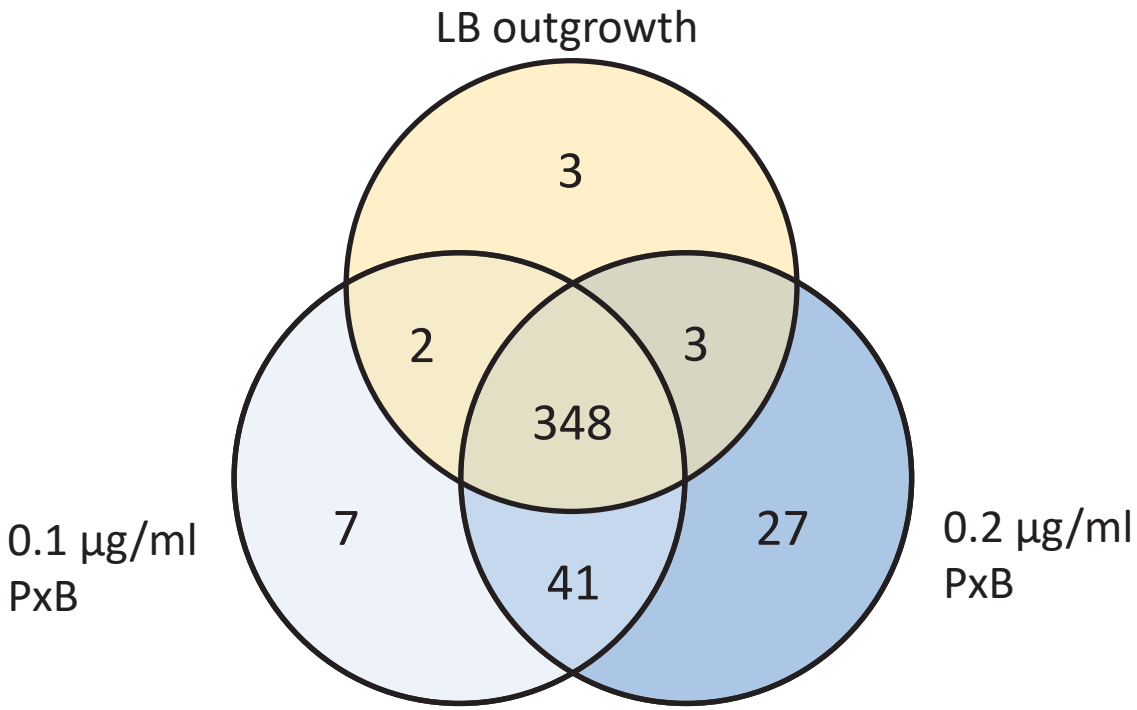

Figure 5.7 Comparison of conditionally essential genes

The identified essential genes for each of the polymyxin B datasets were compared to the LB outgrowth data (OD_{600}=1.00) presented in Chapter 3: LB (yellow), 0.1 µg/ml polymyxin B (pale blue), 0.2 µg/ml polymyxin B (darker blue). The largest group of genes in common between all three datasets are the 348 core essential genes. 41 genes were identified in both polymyxin screens, but a higher concentration of 0.2 µg/ml polymyxin B identified an additional 27 genes required for growth in the presence of polymyxin B. There were 3 genes that were only essential in the LB outgrowth dataset, and 2 and 3 genes that overlapped between the LB outgrowth and one concentration of polymyxin B (0.1 and 0.2 µg/ml respectively).

outgrowth plus one other dataset, also share the same insertion profiles between all three datasets, and are also thought to be falsely classified. Finally, there are 41 genes that are identified as conditionally essential for low level exposure to polymyxin B, and a further 27 genes are reported as conditionally essential genes at an increased concentration of polymyxin B. A full list of conditionally-essential genes and their functions can be found in Table 5.3.

5.2.7 Classification of conditionally essential genes

To determine whether there is any functional enrichment among the conditional essential genes, the relative abundance of cluster of orthologous group (COG) categories were calculated for two lists of genes: WT non-essential genes and the 68 polymyxin B conditionally-essential genes (identified in the 0.2 µg/ml dataset; Appendix 5.1). COG groups are assigned to proteins of known function, to classify proteins of related function (Tatusov *et al.* 2000). FASTA sequences were downloaded from NCBI (CP009273.1) and COG categories determined using the program eggNOG-mapper and the DIAMOND mapping mode (v4.5.1; Huerta-Cepas *et al.* 2017). The relative abundance of each COG category for each dataset is shown in figure 5.8. The polymyxin dataset is enriched for genes in COG groups D, J, M, O and T relative to the total COG data for the whole genome; the most notable are COG groups M and T, which comprise proteins involved in cell envelope biogenesis and signal transduction mechanisms, respectively (Fig. 5.8).

Upon closer inspection, the functional enrichment of COG categories D, J and O were false positive results. All except 2 genes within these lists have the same insertion profiles within the LB only outgrowth as the polymyxin dataset. The false classification of these genes as essential in the polymyxin B dataset but not the LB

Table 5.3 Genes required for growth in LB supplemented with polymyxin B

Gene	Function[a]
ackA	Acetate kinase and propionate kinase
arcA	ArcA transcriptional regulator; activated by ArcB
bamB	Part of the OMP assembly complex
cydC	Glutathione/L-cysteine exporter, part of cytochrome *d* complex
dapF	Diaminopimelate epimerase, catalyses the penultimate step in the biosynthesis of lysine
dksA	Binds RNA polymerase and augments the effect of ppGpp on transcription initiation. Involved in translational regulation of RpoS.
dnaK	Chaperone protein
dnaQ	DNA polymerase III subunit ε
fdx	Reduced ferredoxin
ftsB	Cell division protein, required for localisation of FtsW and FtsI to the Z ring
ftsL	Cell division protein, required for localisation of FtsI to the Z ring
galE	UDP-glucose 4-epimerase, part of galactose catabolism pathway
galU	UTP-glucose-1-phosphate uridylyltransferase. GalU catalyzes the synthesis of UDP-D-glucose from UTP and alpha-D-glucose 1-phosphate
glnD	PII uridylyltransferase
gmhB	Encodes phosphatase of ADP-heptose biosynthesis pathway
grpE	Nucleotide exchange factor in the DnaK-DnaJ-GrpE chaperone system
hldD	The last enzyme in the ADP-heptose synthesis pathway, precursor of core LPS
hldE	Bifunctional: kinase and adenyltransferase
holC	DNA polymerase III subunit χ
iraM	Anti-adaptor protein IraM, inhibitor of σ^S proteolysis
lpp	Murein lipoprotein, stabilises the bacterial cell envelope by tethering the outer membrane to the peptidoglycan layer
mreC	Inner membrane protein required for maintaining the bacterial rod shape
mtn	Catalyses glycosidic bond cleavage
ndh	NADH:quinone oxidoreductase II, catalyses the transfer of electrons from NADH to the quinone pool in the cytoplasmic membrane
nhaA	Na^+:H^+ antiporter, dimer within the membrane
nudB	Dihydroneopterin triphosphate pyrophosphohydrolase
pgm	Phosphoglucomutase, involved in the breakdown of glycogen and metabolism of galactose and maltose
priA	Phosphoglucomutase, involved in the breakdown of glycogen and metabolism of galactose and maltose
pta	Phosphate acetyltransferase, catalyses the reversible conversion between acetyl-CoA and acetylphosphate
ratA	Ribosome association toxin RatA, inhibits translation through binding free 50S to prevent 70S ribosome formation
rbfA	30S ribosome binding factor required for processing of the 16S rRNA
rcsB	RcsB DNA-binding transcriptional activator, part of the RcsF/RcsC/RcsD/RcsA-RcsB phosphorelay system involved in regulating capsular polysaccharide

rcsC	Sensory histidine kinase
rcsD	RcsD phosphotransferase
rimP	Ribosome maturation factor RimP, assists in the maturation of the 30S subunit
rnc	RNase III
rne	Rnase E, involved in the processing and cleavage of RNA
rnhA	Rnase HI, cleaves RNA in RNA-DNA complexes
rnpA	Rnase P protein component, processes tRNA precursor molecules and the 4.5 S RNA precursor
rpiA	Ribose-5-phosphate isomerase A, part of the pentose phosphate pathway
rpmE	50S ribosomal subunit protein L31
rpmF	50S ribosomal subunit protein L32
rpoS	RNA polymerase, sigma S factor (sigma 38)
rseA	anti-sigma factor that inhibits σ^E, until degraded by DegS upon extracytoplasmic stress.
sapD	Putrescine ABC exporter ATP binding protein
secF	Sec translocon accessory complex subunit
smpB	Small protein B, a component of the trans-translation process which marks proteins for degradation that have stalled during translation
surA	A chaperone involved in the correct folding and assembly of OMPs
tktA	Transketolase I
tpr	Protamine-like protein
trpL	*trp* operon leader
tusD	Sulfurtransferase complex subunit
waaB	Adds galactose from UDP-galactose to the first glucose residue (Glcl) of the LPS outer core
waaC	ADP-heptose:LPS heptosyltransferase I - transfers the first heptose sugar onto the Kdo2 moiety of the lipopolysaccharide inner core.
waaF	ADP-heptose:LPS heptosyltransferase 2, transfers the seconds heptose sugar to heptosyl-Kdo2 moiety of LPS inner core.
waaR (waaO)	Glucosyl transferase, adds the second glucose (GlcII) to the first glucose (GlcI) of the LPS outer core
waaP	LPS kinase, required for the addition of phosphate to O-4 of the heptose I residue in the LPS core
ychF	Ribosome-binding ATPase
yddK	Leucine-rich repeat domain-containing protein
ydfA	Qin prophage; putative protein
ydfK	Qin prophage; cold shock protein YdfK
ygfZ	Folate-binding protein of unknown function
ymiB	Putative protein
ynaE	Rac prophage; cold shock protein
yobI	Uncharacterised protein
yqcG	Cell envelope stress response protein
yqeJ	Putative protein
yqeL	Uncharacterised protwin

[a]Functions taken from the Ecocyc website: https://ecocyc.org/

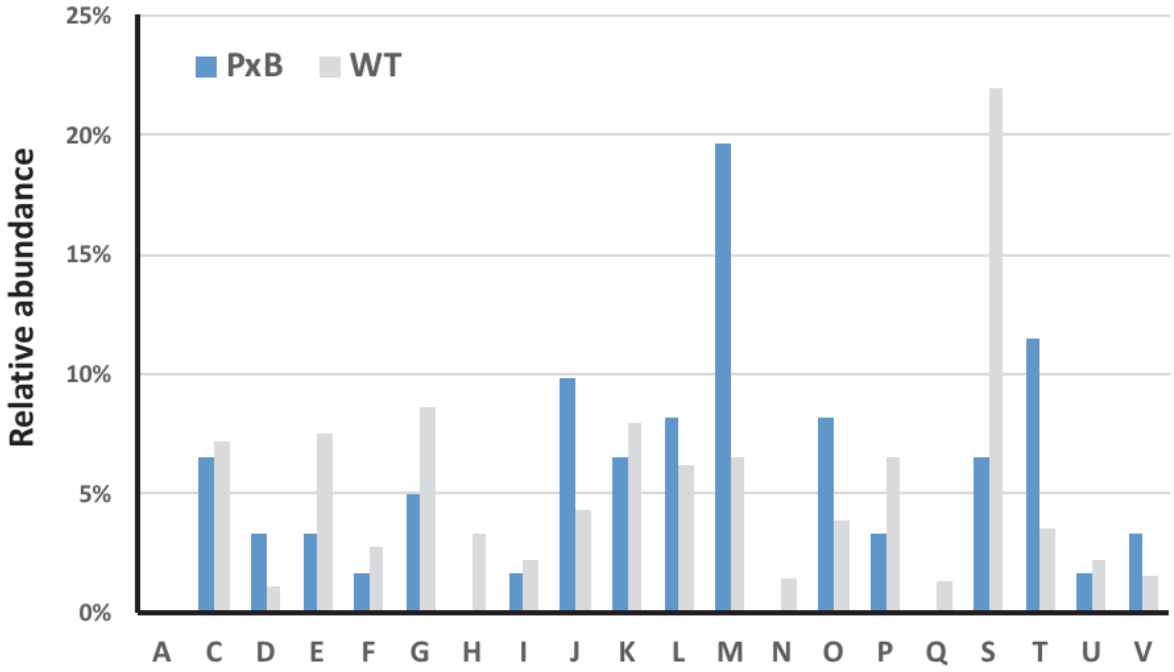

Figure 5.8 Comparison of the relative abundance of genes within COG categories

Cluster of orthologous groups (COG) categories were identified for each of the 68 conditionally essential genes identified in the TraDIS screen with 0.2 µg/ml polymyxin B. The relative abundance of genes grouped by COG categories from the polymyxin dataset (blue) were compared to the total COG classification abundance in *E. coli* BW25113 (grey). There is a marked increase in conditional essential genes in the polymyxin dataset in COG group 'M', which comprises proteins involved in cell wall, membrane and envelope biogenesis.

outgrowth dataset is one of the limitations of predicting gene essentiality based on the relative frequency of insertion per CDSs within a dataset. The subsequent COG analysis is only as good as the input data and would suggest alternate, more scrupulous, data analysis methods might be more appropriate for identifying conditionally essential genes. Furthermore, effective COG analysis is dependent upon adequate annotation of the reference genome. However, inspection of the genes within COG groups M and T revealed conditionally essential genes that were examined further in the subsequent sections.

5.2.8 Characterization of LPS mutants sensitive to polymyxin B

COG analysis revealed that genes involved in cell envelope biogenesis were enriched among the list of conditionally essential genes. As LPS is the outermost barrier of the envelope and polymyxin B is reported to bind LPS, genes involved in LPS biosynthesis were investigated further (Mares *et al.* 2009). A number of genes within the *waa* operons had significantly fewer insertions compared to the LB-outgrowth data, specifically the genes *waaR*, *waaB*, *waaP*, *waaG*, *waaF* and *waaC* (Fig. 5.9A). The enzymes that these genes code for were mapped onto the structure of LPS. Most side branch components appear dispensable, with the exception of galactose and the first phosphate added by WaaP.

Consistent with these data, the last three genes in the pathway for D-glucose and D-galactose synthesis (*pgm*, *galU* and *galE*), and the committed steps of the D-manno-heptose biosynthetic pathway were also identified as conditionally essential (Fig. 5.10). This result correlates well with the enzymes involved in LPS biosynthesis that are identified as conditionally essential, as inner-core heptose is ligated to

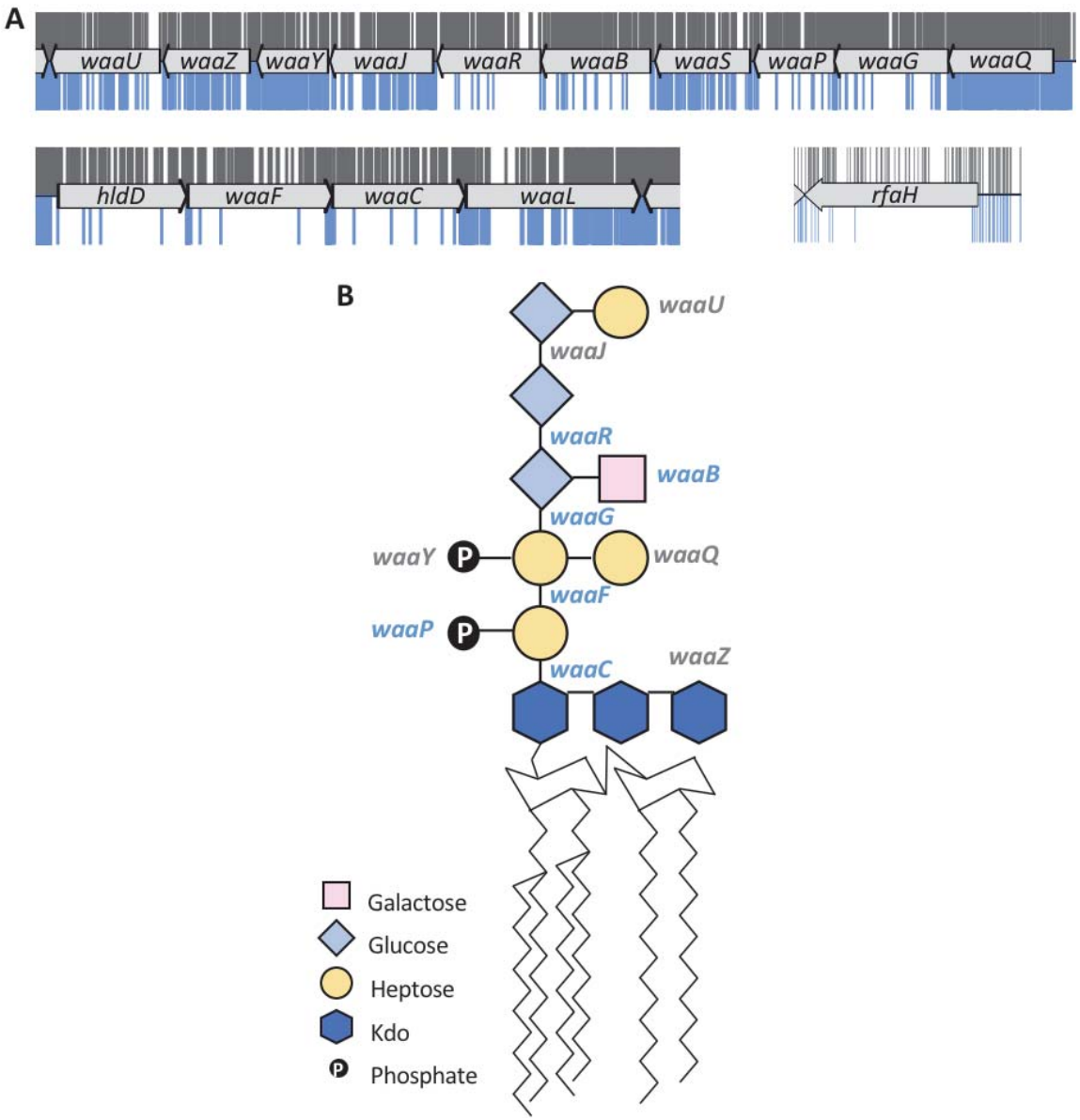

Figure 5.9 Inspection of LPS mutants sensitive to polymyxin B

(A) TraDIS data for growth in LB (grey) or LB + 0.2 μg/ml polymyxin (blue) were plotted above and below the gene track, respectively. Insertion sites are represented by vertical bars and mapped reads are capped at a frequency of 1. Insertions within genes required for LPS core biosynthesis are significantly underrepresented. (B) A schematic of the structure of LPS. LPS biosynthesis enzymes are indicated next to the linkage they form (central column) or component they ligate (side branches). Conditionally essential enzymes are labelled in blue.

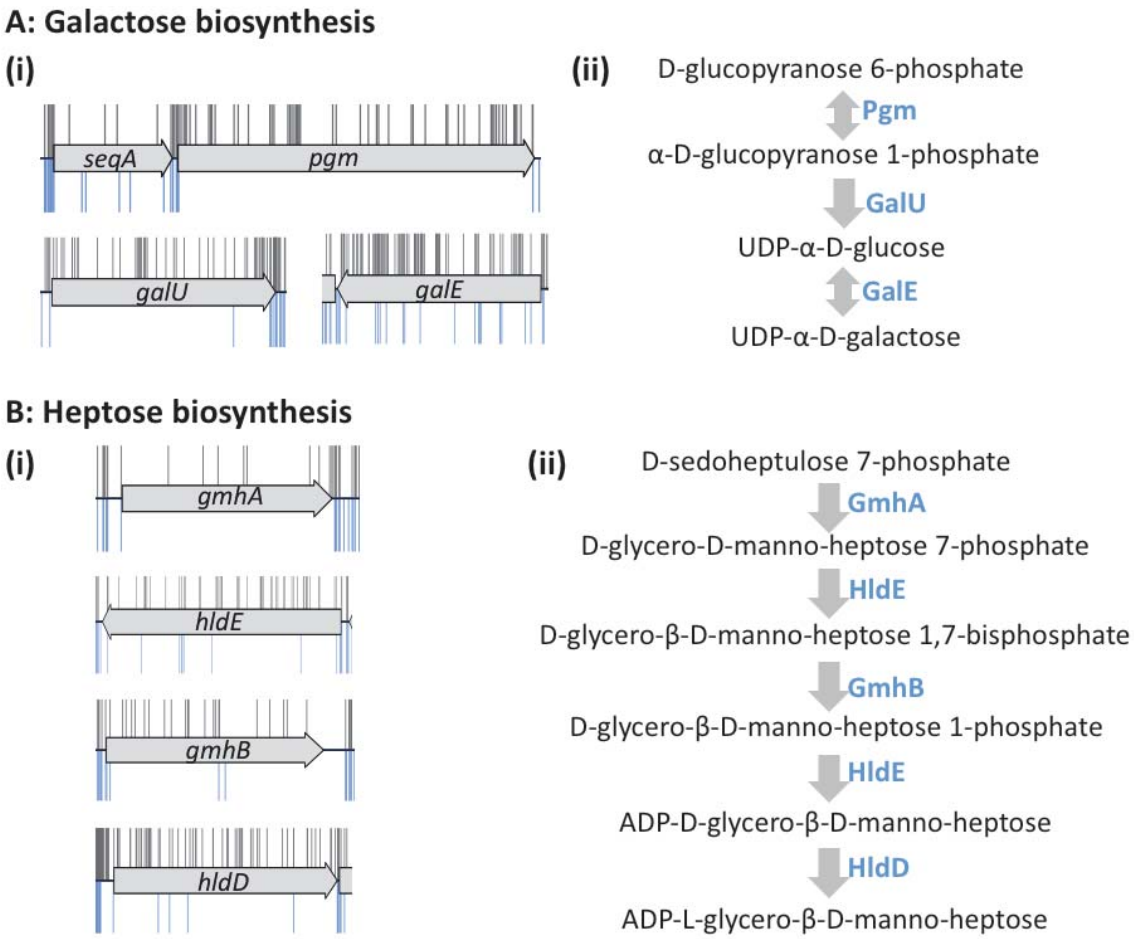

Figure 5.10 Committed steps of the galactose and heptose biosynthetic pathways

TraDIS data for growth in LB (grey) or LB + 0.2 μg/ml polymyxin (blue) were plotted above and below the gene track, respectively. Insertion sites are represented by vertical bars and mapped reads are capped at a frequency of 1. (A) The committed steps of the D-galactose biosynthetic pathway were conditionally essential for growth in the presence of polymyxin B. (B) The committed steps of the heptose biosynthetic pathway were conditionally essential for growth in the presence of polymyxin B (blue).

Kdo-lipid A by WaaC and WaaF, glucose is added by WaaG, and Galactose is added by WaaB.

Together, these data suggest that the backbone of the LPS inner core is required for survival in the presence of sub-inhibitory concentrations of polymyxin B, as is the first glucose of the outer core and the associated galactose molecule. It has already been reported in the literature that LPS mutants are more sensitive to polymyxin B (Fernández *et al.* 2013), but these data reveal the specific components of the LPS structure that are important for maintaining barrier function in the presence of polymyxin B.

5.2.9 Comparison of mutants with defects in OMP trafficking

The OM is the first barrier to any antimicrobial agents, and, in addition to LPS, contains OMPs and lipoproteins. Having established that transposon mutants defective in LPS synthesis are more sensitive to polymyxin B, the data were then scrutinised to determine whether any defects in OMP transport or assembly also result in increased sensitivity to polymyxin B. The BAM complex (β-barrel assembly machinery) catalyses folding and insertion of OMPs into the outer membrane (Han *et al.* 2016; Knowles *et al.* 2009). It consists of 5 proteins: BamA, BamB, BamC, BamD and BamE. BamA and BamD are essential and so cannot be assessed in a TraDIS whole-genome screen, however, the accessory lipoproteins BamB, BamC and BamE are all non-essential. Inspection of the TraDIS data reveals that BamB and BamE are conditionally essential for growth in the presence of polymyxin B, but BamC is not (Fig. 5.11 A and B). The insertion profiles of the *bamB* and *bamE* genes clearly show a stark loss of insertions following outgrowth with polymyxin B, while the insertion profile of *bamC* resembles that of growth in the absence of polymyxin B (Fig. 5.11A).

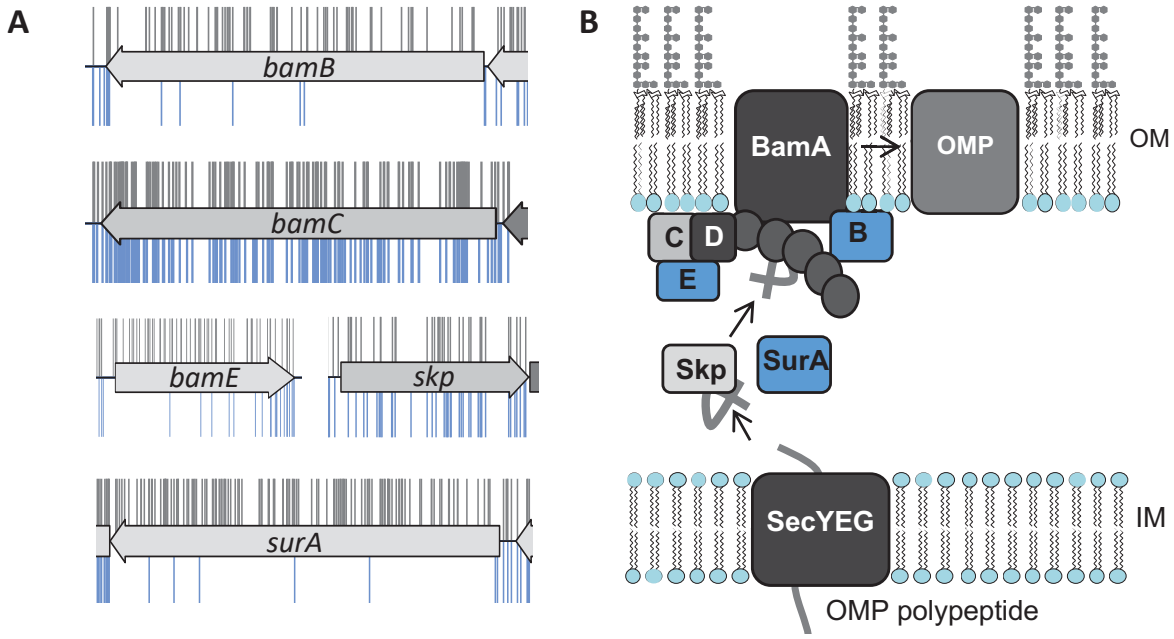

Figure 5.11 Components of the OMP trafficking and assembly pathways required for growth in the presence of sub-inhibitory concentrations of polymyxin B

(A) TraDIS data for growth in LB (grey) or LB supplemented with 0.2 µg/ml polymyxin (blue) were plotted above and below the gene track, respectively. Insertion sites are represented by vertical bars and mapped reads are capped at a frequency of 1. (B) Schematic of protein translocation and trafficking to the BAM complex. Essential proteins are shown in dark grey, conditionally essential proteins are shown in blue and dispensable proteins are shown in light grey.

Furthermore, β-barrel proteins destined for the OM are transported by the chaperones SurA and Skp (Sklar *et al.* 2007). In the presence of polymyxin B *surA* is essential for growth but *skp* is not (Fig. 5.11A).

These results are likely all due to the role of LptD. LptD is a β-barrel in the OM that inserts LPS into the OM (Wu *et al.* 2006). LptD assembly into the OM is mediated by the BAM complex in conjunction with the lipoprotein LptE (Lee *et al.* 2016). Furthermore, LptD is trafficked to the BAM complex by SurA (Vertommen *et al.* 2009). In a *surA* deletion strain the amount of OM LptD is 10% of the amount of OM LptD detected in the parent WT strain (Vertommen *et al.* 2009). LptD cannot be deleted because it is essential (Baba *et al.* 2006; Goodall *et al.* 2018; Braun & Silhavy 2002), an LptD depletion strain filaments and lyses (Braun & Silhavy 2002), and a known LptD mutant, 'imp4213', has an increased cell permeability (Eggert *et al.* 2001; Ruiz *et al.* 2005). Therefore, it seems likely that the observed polymyxin B sensitivity of bam and SurA mutants is due to defective LptD assembly. This might suggest that the lipoproteins BamB and BamE, but not BamC are required for proper LptD assembly.

Defects in LPS assembly and trafficking to the OM are reported to result in an increase in cell permeability attributed to an increase in surface exposed phospholipids (Nikaido 2005). Secondly, polymyxin B is reported to have a higher affinity for phospholipids than LPS (Domingues *et al.* 2012). Therefore, it is possible that the increased sensitivity of these mutants to polymyxin B is due to increased exposure of surface phospholipids, which might facilitate increased uptake of polymyxin B molecules.

5.2.10 Validation of data from TraDIS experiments

To validate some of the TraDIS results, individual gene disruption mutants were constructed for the genes *bamB*, *galU*, *rfaP* and *surA*. Kanamycin resistant mutants from the Keio library were transferred into strain BW25113 by P1 transduction. Strains were grown overnight, and then diluted to an OD_{600} = 0.10. From this initial starting inoculum, cells were 2-fold serially diluted (Fig. 5.12A). A volume of 2 µl was spotted onto agar plates supplemented with or without polymyxin B at concentrations of 0.1 or 0.2 µg/ml and the plates were incubated overnight at 37°C. Solid medium was used to ensure the polymyxin B concentration was equivalent among all strains, allowing a direct comparison between strains. In addition, 2 µl of cells with an OD_{600} = 0.1 were inoculated onto a gradient plate, across the width of the plate, with a concentration of polymyxin B ranging from ~0-1 mg/ml (Fig. 5.12B).

The differences were slight, but in both variants of the experiment, the wildtype strain was less susceptible than the mutants to polymyxin B. The effect was most apparent at 0.1 µg/ml polymyxin B. Among these mutants, the *surA* deletion mutant is the most sensitive to polymyxin B. While these mutants were screened on solid medium, and the initial TraDIS experiment involved the growth and identification of mutants in broth medium, these data still validate the initial findings that these mutants are more sensitive to polymyxin B than the WT strain.

5.2.11 Limitations of the TraDIS analysis

A limitation of the bi-modal classification of essential genes method is that gene classifications are calculated relative to the rest of the population. For example, a gene with a low number of insertions, but sufficient insertions that it does not meet the criteria to be classified as essential in the LB outgrowth data, might be re-classified

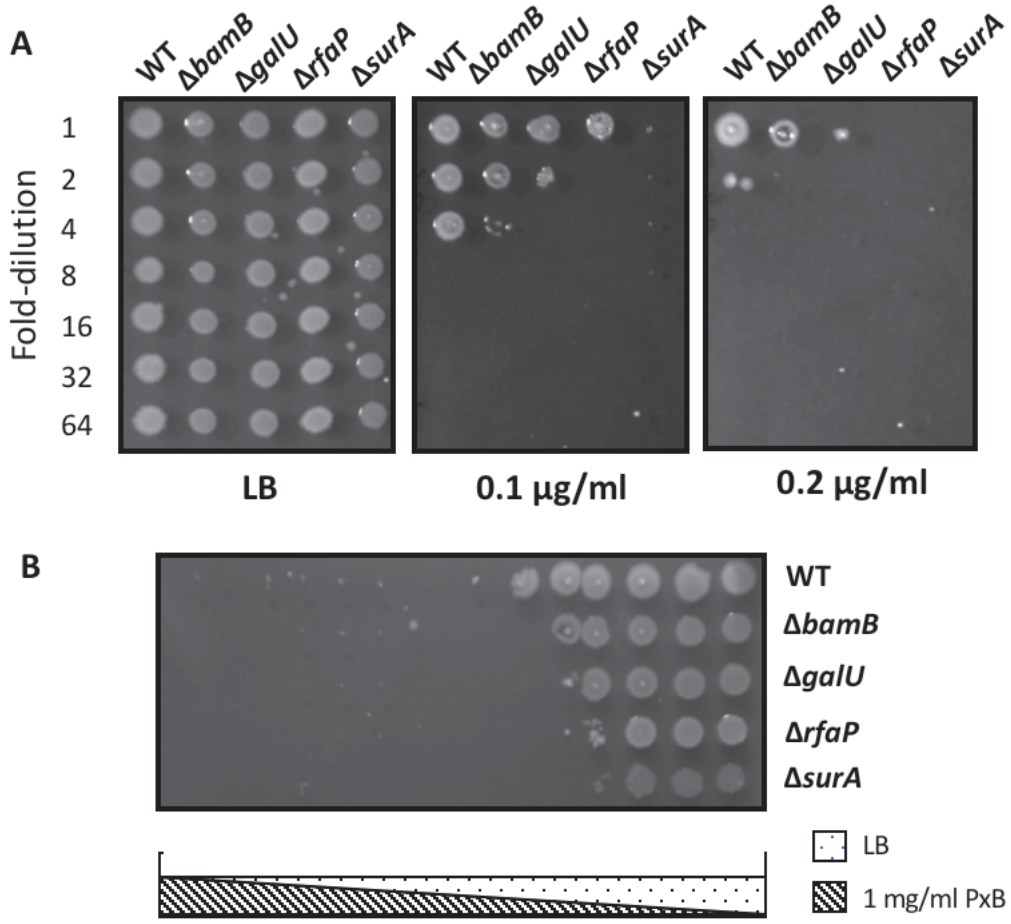

Figure 5.12 Validation of polymyxin B-sensitive mutants identified by TraDIS

Strains (from left to right) were BW25113 (WT), and BW25113 single gene deletions: Δ*bamB*, Δ*galU*, Δ*rfaP* and Δ*surA* grown overnight and diluted to OD_{600}=1.00. (A) 2 µl of 2-fold serially-diluted strains (starting from OD_{600} = 0.1) were inoculated onto agar plates with and without polymyxin B. (B) 2 µl (at OD_{600} = 0.1) of each strain was inoculated across the gradient plate. PxB, polymyxin B.

as essential following outgrowth with polymyxin B (despite an identical insertion) as this dataset has a larger number of total essential genes, resulting in a shift of the overall essential mode to the right (on the x-axis of an IIS frequency plot). For example, the genes *rbfA* and *rnpA* are classified as essential in the polymyxin 0.2 µg/ml dataset, but are classified as 'unclear' in the LB outgrowth dataset, despite having identical or very similar insertion index scores (Fig. 5.13). Discrepancies like these are only identified following manual inspection.

Another limitation of this analysis method is that conditional-essential judgements are based on the absence or presence of insertions and does not consider mutant abundance. For example, there might be some mutants that are more sensitive to polymyxin B than the parent strain, but as the library has only gone through approximately 5 generations of growth these mutants are still present within the total population and will not be identified by quantification of IIS alone. For example, the genes *yhcB* and *ydbH* are visibly less well represented by mutants but are classified as non-essential when quantifying genes by number of insertions (Fig. 5.14). For genes like these a more appropriate analysis method would be a sliding-window approach that quantifies read abundance relative to neighbouring insertion sites.

5.3 Conclusion

The data presented in this chapter confirms that TraDIS is a tool that can be used to identify genes required for survival in a given condition. Especially remarkable is that TraDIS can identify sensitive mutants even when the there is only a 2-fold difference in sensitivity between the wildtype strain and mutant. The sensitivity of this method

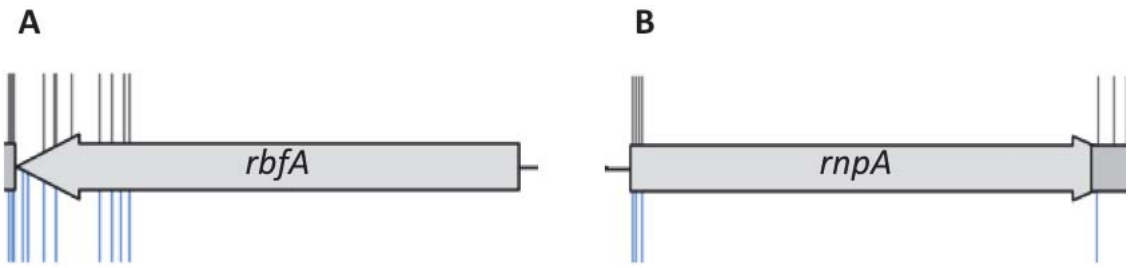

Figure 5.13 Genes with comparable insertion index scores but different essentiality classifications

TraDIS data for growth in LB (grey) or LB supplemented with 0.2 µg/ml polymyxin B (blue) were plotted above and below the gene track, respectively. Insertion data are shown as peaks above and below the gene track and capped at a frequency of 1 to represent unique insertion sites. Both of the genes *rbfA* and *rnpA* were classified as essential in the polymyxin datasets but not LB outgrowth.

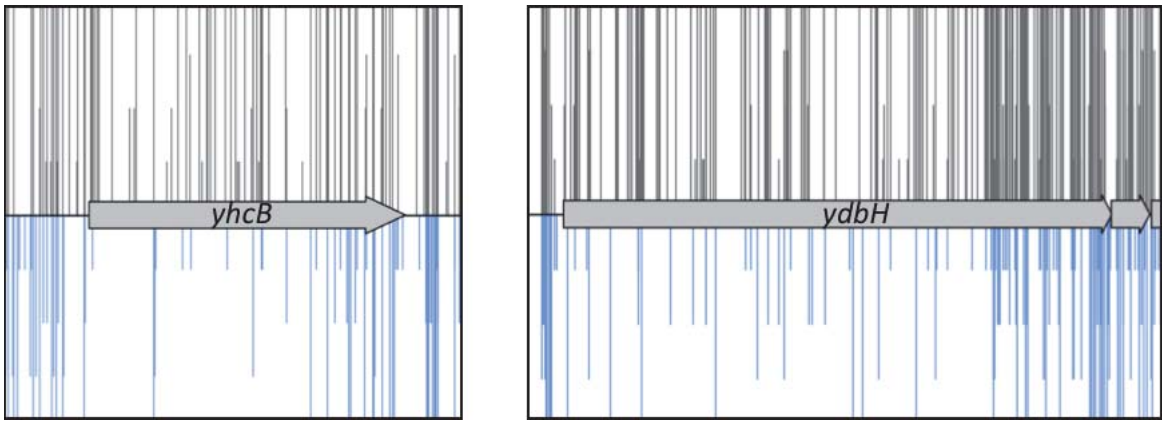

Figure 5.14 Genes with low insertion index scores that were not statistically classified as conditionally essential

The BW25113 chloramphenicol transposon library was grown in LB supplemented with and without sub-MIC concentrations of polymyxin B, shown above (grey) and below (blue) the gene track respectively. The insertion frequency is capped at 5. Both *yhcB* and *ydbH* visibly display a decrease in insertion frequency following outgrowth in LB supplemented with 0.2 µg/ml polymyxin B.

enables the identification of specific components of macromolecular structures required for antimicrobial resistance at the molecular level of resolution. The importance of LPS, specifically the integrity of LPS crosslinking, was already known to be important for withstanding polymyxin treatment but the significance of galactose is a new result and requires further validation (Trimble *et al.* 2016). This is an important strength of the method as the ability to identify specific components of LPS that confer intrinsic resistance to polymyxin B might aid the development of more effective methods of treatment. For example, co-treatment with an adjuvant that targets specific components of LPS, or cleaves specific linkages. This might then enable patient treatment with lower doses of polymyxin as polymyxins are notoriously toxic to humans due to indiscriminate disruption of the eukaryotic plasma membrane (Abdelraouf *et al.* 2012). Lower clinical uses of polymyxin might then reduce the selective pressure exerted on bacteria and slow the onset of polymyxin resistant strains.

A limitation of this method, however, is that it is impossible to screen essential genes, and some non-essential genes, that are not represented within the transposon library. Furthermore, the quality of the output data is influenced by the analysis methods used. Either amendments to the growth protocol might be required to further distinguish between sensitive mutants and the WT, such as prolonged outgrowth or passaging of the library, or additional analysis methods that take into account the frequency with which each mutant is observed. Such analysis methods might include a Hidden Markov Model approach such as the analysis published by DeJesus & Loerger (2013), which applies a sliding window across data to identify variations in insertion frequency relative to the neighbouring genomic regions.

However, inspection by eye did reveal 2 genes of unknown function with putative roles in maintaining intrinsic resistance to polymyxin B: *yhcB* and *ydbH*.

Chapter 6

A Possible Role for YhcB in Envelope Biogenesis

6.1 Introduction

Genes that, when disrupted, result in a fitness cost in response to an antibiotic might code for products that are good candidates as antimicrobial targets. In the previous chapter, the gene *yhcB* was identified in a screen of BW25113 transposon mutants grown in LB supplemented with sub-MIC concentrations of polymyxin B. In addition, a *yhcB* transposon mutant of *K. pneumoniae* was previously reported to be sensitive to colistin (Jana *et al.* 2017). These data suggest that *yhcB* might have role in maintaining envelope integrity.

To date there are only 3 published papers about *yhcB*. YhcB was initially classified as part of the cytochrome *bd* ubiquinol oxidase (Stenberg *et al.* 2005). However, this classification was later challenged when Mogi *et al.* (2006) showed that cytochrome *bd* function and assembly are both unaltered in a *yhcB* deletion strain compared to the parent. Furthermore, *yhcB* is conserved only in gamma-proteobacteria, unlike *cydAB*, which are widely distributed. The authors concluded that YhcB is likely not part of the cytochrome *bd* complex (Mogi *et al.* 2006), but is part of either a larger protein complex or a multimeric complex that migrates to the same position as, or near to, cytochrome *bd* in a one-dimensional blue-native PAGE gel.

More recently, *yhcB* was reported to have a synthetic lethal relationship with *rodZ*, and the protein was reported to interact with RodZ, MurG, MreC, MreD, RodA, YciS (LapA) and itself in bacterial two-hybrid (BACTH) assays (Li *et al.* 2012). All of these proteins are involved in cell wall synthesis, determining cell shape or LPS assembly. As *yhcB* was identified in a screen designed to highlight mutants with envelope defects, YhcB might have a role in the synthesis or maintenance of the cell

envelope. The aim of this chapter was to characterise *yhcB*, in an attempt to understand the role of this gene of unknown function.

6.2 Results

6.2.1 The conservation and structure of yhcB

In *E. coli,* the gene *yhcB* is divergently transcribed from *zapE* and upstream of *degQ* and *degS*, and encodes a protein of 132 amino acids (Fig. 6.1). There is one reported promoter for *yhcB*, although sequence inspection and dRNA-seq data suggest there are two overlapping promoters (Mendoza-Vargas *et al.* 2009; Keseler *et al.* 2017; Thomason *et al.* 2015). YhcB is reportedly conserved in gamma-proteobacteria (Mogi *et al.* 2006). Within the pfam database, the majority of the *yhcB* coding sequence (residues 6-127) correspond with a domain of unknown function: DUF1043. The pfam database reports that DUF1043 is present in a total of 368 species, 364 of which are proteobacteria (354 gamma-proteobacteria and 10 beta-proteobacteria, respectively); the distribution of these species is presented in Fig. 6.2 (Finn *et al.* 2014; date accessed: 25/09/18). To determine whether the sequence of *yhcB* is conserved among species, a multisequence alignment file was created for the amino acid sequence of YhcB from 150 different species using the online program ConSurf (default settings; Ashkenazy et al. 2016). The conserved residues of *yhcB* are presented in Fig. 6.3. Notably, *yhcB* contains two highly conserved histidine residues at positions 55 and 76 within a section of 7-residue periodical repeats, indicative of a helical structure, as well as two highly conserved domains at residues 93-95 and 114-118. However, further literature searches have not identified functional roles for combined 'NPF' or 'PRDY' domains.

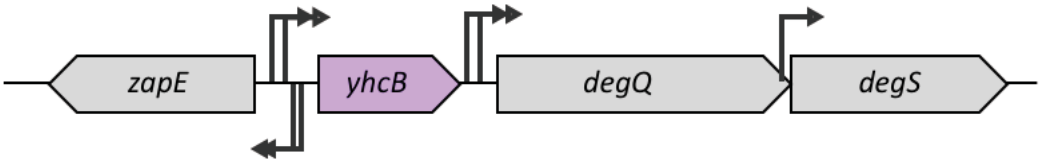

Figure 6.1 The genetic neighbourhood of *yhcB* in *E. coli K-12*

(A) The gene *yhcB* is positioned between *zapE* and *degQ* on the *E. coli* BW25113 chromosome. Promoters are represented by bent arrows. The genes *zapE* and *yhcB* are divergently transcribed and have overlapping 5′ UTRs.

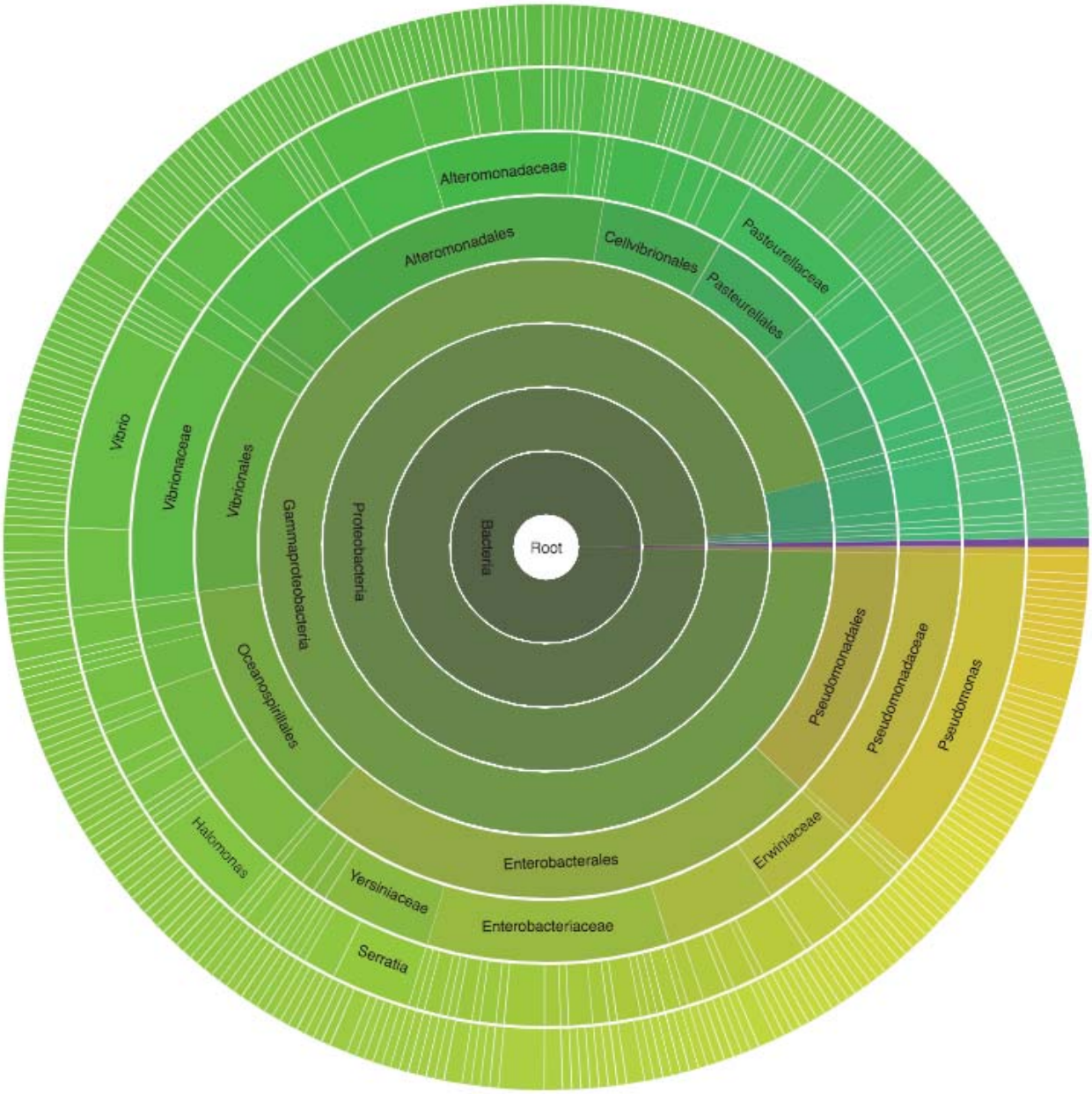

Figure 6.2 Conservation of the DUF1043 domain

Within the pfam database, residues 6-127 of the 132 amino acid protein YhcB are annotated to correspond with a domain of unknown function DUF1043. The DUF1043 domain is conserved within gamma-proteobacteria. The 366 bacterial species containing a DUF1043 domain are shown in green-yellow and are clustered hierarchically. The single eukaryotic species, *Rhodnius prolixus*, found to contain a DUF1043 domain is shown in purple.

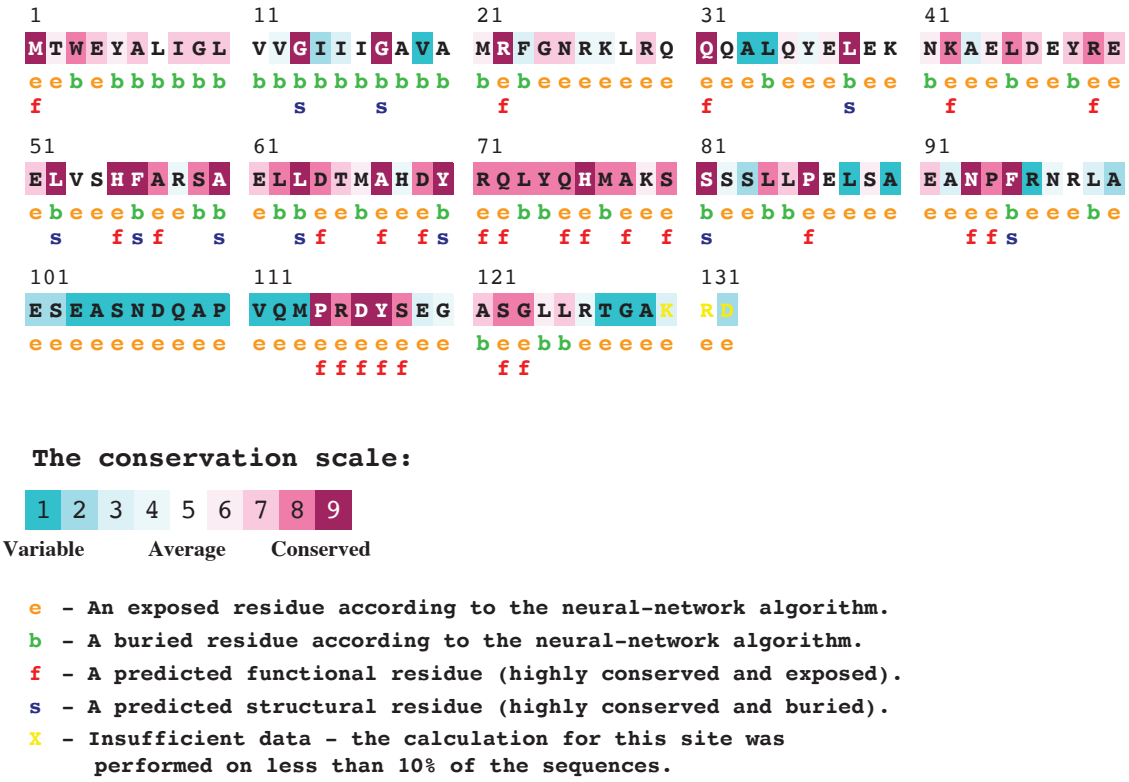

Figure 6.3 Conserved residues of YhcB

The conserved residues of YhcB identified from a multi sequence alignment file containing the sequence of YhcB taken from 150 species. The sequence presented is taken from *E. coli* BW25113. Conservation of residues is indicated by a blue-red scale, highly conserved residues are shown in dark red. Exposure of a residue is predicted based on the residue properties and neighbouring residues. Conserved and buried residues are predicted to have a structural role, conserved and exposed residues are predicted to have a functional role.

The structure of YhcB has not been reported, however, YhcB is known to be an inner membrane protein (Maddalo *et al.* 2011). To understand the possible role of YhcB within the cell, the secondary structure of YhcB in BW25113 was predicted using the program PSIPRED, where each individual amino acid is assigned to a predicted secondary structure. The predicted structure of YhcB is a C-terminal single alpha-helix transmembrane domain, with an additional, longer, cytoplasmic alpha helical domain followed by a coiled or unstructured N-terminal domain, consistent with previously reported data (Fig. 6.4; Mogi *et al.* 2006). These data were combined to produce a schematic of the overall YhcB protein structure, presented in figure 6.5. Finally, the helical cytoplasmic domain was predicted (using Phyre2) to form a coiled-coil structure that closely resembles a HAMP domain (with a confidence score of 89.9/100; Kelley *et al.* 2015), suggesting that YhcB might have a role in signal transduction.

6.2.2 Validation of a ΔyhcB mutant cell envelope defect

A gene deletion mutant was constructed to validate the TraDIS findings. The *yhcB* gene was replaced via P1 transduction of the kanamycin resistance cassette from the *yhcB* mutant in the Keio library collection of mutants in BW25113. The kanamycin cassette was then excised by Flp recombination of the FRT sites within the cassette leaving a 102 nt scar (Fig. 6.6). The successful construction of mutants was confirmed by PCR (Fig. 6.6B).

The antibiotic polymyxin B is notoriously difficult to work with as it is known to bind to plastic, making pipetting and long-term storage of stocks problematic (Bakthavatchalam *et al.* 2018). The activity of polymyxin B is also affected by cations within the media, which can be variable in an undefined medium such as LB

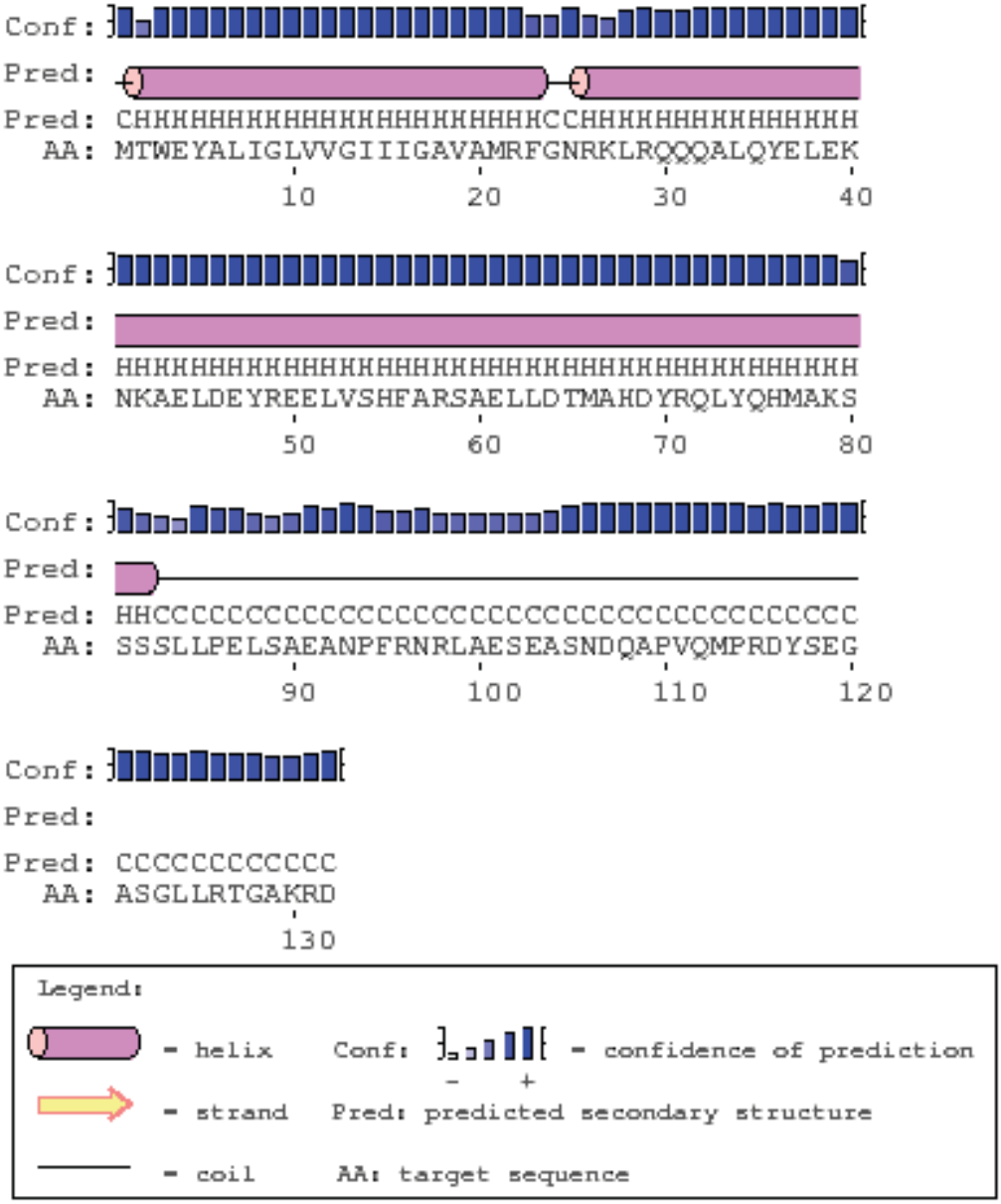

Figure 6.4 The predicted secondary structure of YhcB

The secondary structure of YhcB predicted using the program PSIPRED. Each individual amino acid is assigned to a predicted secondary structure. Alpha-helix structure is indicated by a pink tube, coiled structure is indicated by a black line. The confidence of the prediction is indicated by the height of the blue bar above each amino acid position. YhcB is predicted to have a short (transmembrane) helix, a longer helix and a disordered C-terminal domain.

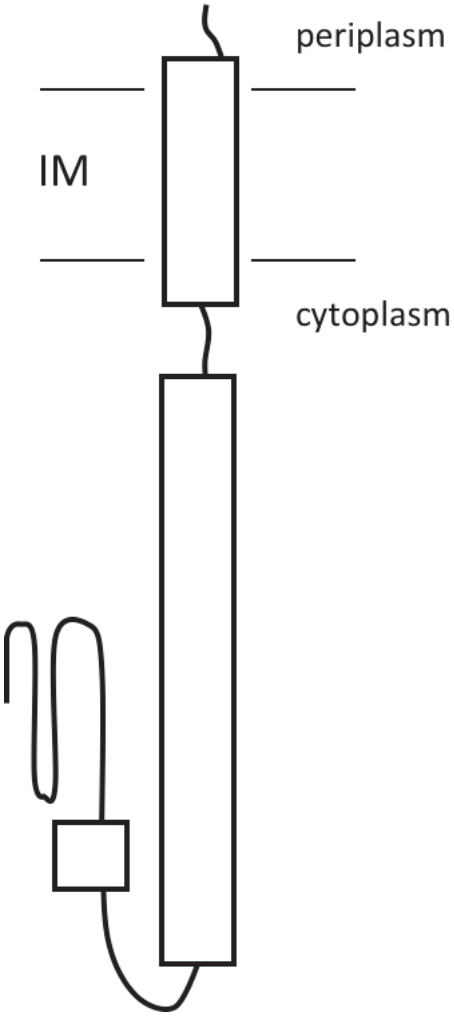

Figure 6.5 Schematic of the predicted secondary structure of YhcB

A schematic of the predicted secondary structure of YhcB adapted from Mogi *et al.* (2006), in conjunction with PSIPRED and ConSurf data. YhcB is an IM protein with a single N-terminal transmembrane pass. YhcB has a cytoplasmic alpha-helix, followed by an unstructured C-terminal domain. The C-terminal end of the protein is predicted to contain conserved residues that might form a 3rd functional domain.

Figure 6.6 Construction of a Δ*yhcB* mutant

(A) The position of *yhcB* relative to *zapE* and *degQ* on the *E. coli* BW25113 chromosome. Primer binding sites are indicated by arrows. (B) The Δ*yhcB* strain was constructed by a two-step method: first the *yhcB* gene was replaced by a kanamycin cassette followed by recombination of the FRT sites within the cassette leaving a 102 nucleotide scar consisting of the *yhcB* start codon, FRT scar, and the last 7 codons of *yhcB* (inclusive of the stop codon). The construction of strains was confirmed by PCR with primers flanking the *yhcB* gene.

(Bakthavatchalam *et al.* 2018; Girardello *et al.* 2012). For this reason, work with polymyxin B was discontinued. Instead, to test the OM-sensitive phenotype, strains were grown on LB agar supplemented with vancomycin, or with SDS and EDTA, two well established screens for cell envelope defects. Microdilution assays were prepared as described in materials and methods. The Δ*yhcB* mutant was 1,000-fold more sensitive to 1% SDS and 0.5 mM EDTA, and did not grow when the medium was supplemented with 1 mM EDTA. The Δ*yhcB* mutant was also sensitive to vancomycin, but this phenotype was less pronounced (Fig. 6.7). These findings support the hypothesis that a *yhcB* mutant has a defective cell envelope. There was also a consistent 10-fold difference in growth on LB agar in a microdilution assay, despite the inoculum being normalised for optical density. This result indicates a division or size defect that might account for a comparable OD but a lower CFU/ml.

6.2.3 Cell morphology of a ΔyhcB mutant

To determine whether a Δ*yhcB* mutant has an observable phenotype that might account for the CFU difference observed previously, strains were grown overnight, sub-cultured, and visualised by microscopy (Fig. 6.8 A and B). The Δ*yhcB* mutant cells were obviously larger, which was quantified by measuring the cell length and width of 50 cells for the parent and mutant strains (Fig. 6.8C). Both the length and the width of Δ*yhcB* mutant cells were significantly larger than the WT (Student's t-test, $p < 0.0001$; Fig. 6.8D). The increase in width was largely uniform, however, the increased length of a Δ*yhcB* mutant was variable, which can be seen in both the microscopy image and the spread of the data for 50 cells (pink, Fig. 6.8 B, C &D). The cell envelope in a Δ*yhcB* mutant was clearly defective under these conditions: the cells were often more sensitive to pressure under the slide and burst, the width was

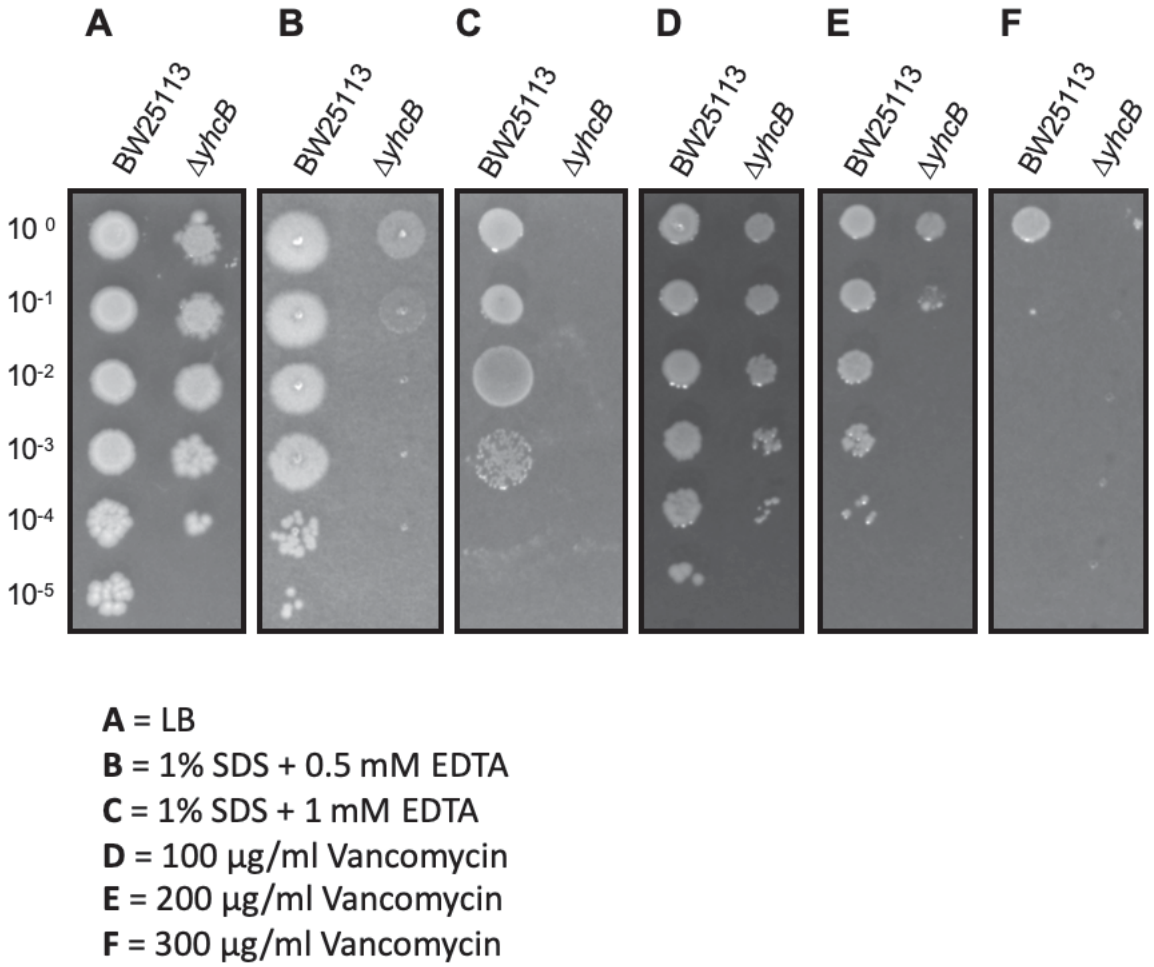

Figure 6.7 Validation of a Δ*yhcB* mutant cell envelope defect

Overnight cultures of the WT strain and Δ*yhcB* strain grown in LB at 37°C were diluted to an OD_{600} of 1.00. These cultures were 10-fold serially diluted in LB. 1.5 μl culture was spotted onto LB (A); or LB supplemented with the following: 1% SDS + 0.5 mM EDTA (B); 1% SDS + 1 mM EDTA (C); 100 μg/ml vancomycin (D); 200 μg/ml vancomycin (E); 300 μg/ml vancomycin (F). The Δ*yhcB* mutants were 100-fold more sensitive to 200 μg/ml vancomycin, and 1,000-fold more sensitive to 1% SDS + 0.5 mM EDTA.

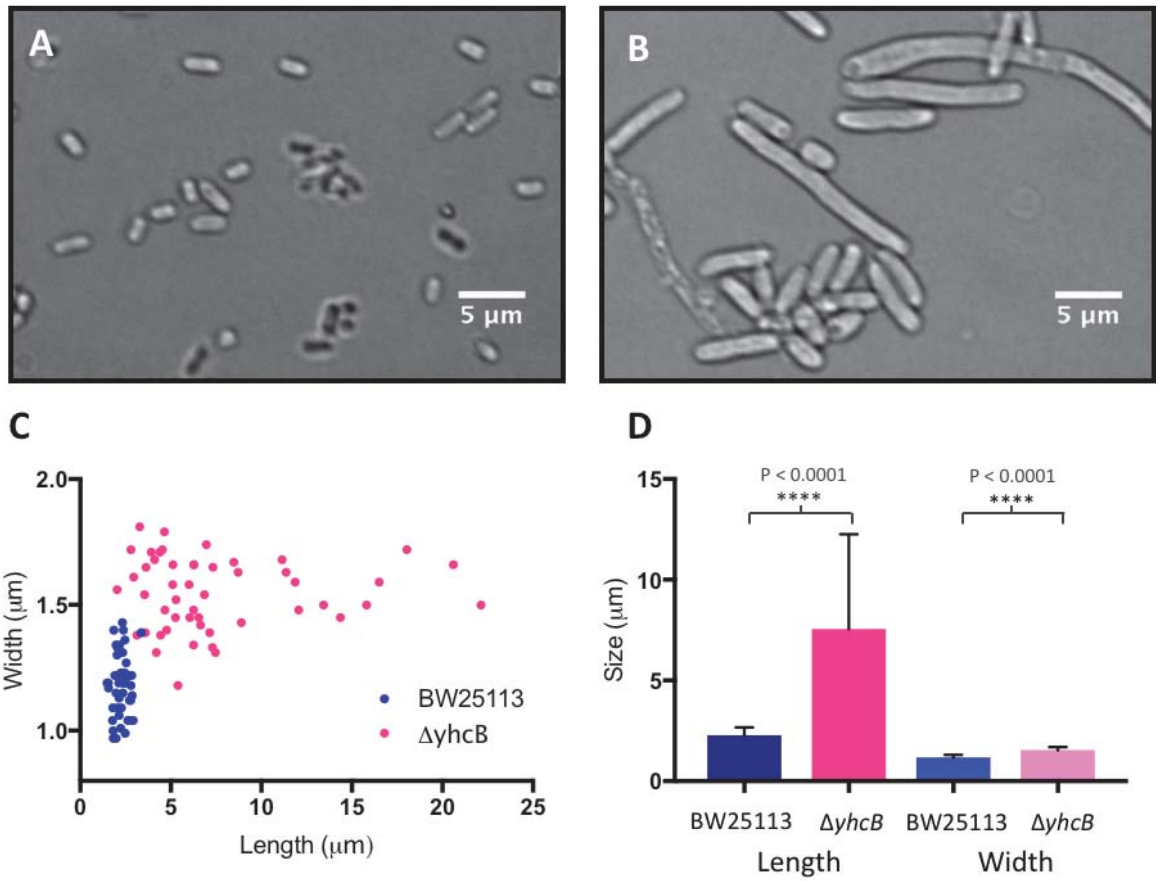

Figure 6.8 The cell morphology of a Δ*yhcB* mutant

The WT (A) and Δ*yhcB* mutant (B) were grown overnight in LB at 37°C. Cells were visualised using a Nikon 90i microscope, a 5 µm bar is shown for scale. The DIC images presented were captured using a 100x objective lens. (C) The length and width of 50 cells were measured for each strain, BW25113 is shown in blue and Δ*yhcB* in pink. (D) The mean length and width of each strain were plotted: WT (blue) and Δ*yhcB* (pink). The error bar represents the standard deviation of the mean. A students t-test with Welch's correction for unequal variance was used to compare strains. Both the length and width of the Δ*yhcB* mutant were significantly larger than the WT strain ($p<0.0001$).

not always consistent along the full length of the cell, occasionally blebs were observed at the septum site, and on at least 2 occasions a mutant was identified with a third branch resulting in a Y-shaped bacterium.

In addition to the observed size defect, the other clear observation was irregular septum positioning. Some cells were captured immediately after division and had an almost normal appearance (despite the enlarged size), while other cells grew to a range of lengths up to ~22 µm with no visible septum or chaining phenotype; these cells were massive, viable, single-cell entities! The variation in cell length is suggestive of a defect in the timing of division. Although it is not clear whether the mis-timed Z-ring is secondary to increased cell volume, or whether the larger cell is a consequence of failed division.

6.2.4 Growth kinetics of a ΔyhcB mutant

The apparent cell cycle defect was explored further by investigating the growth of a Δ*yhcB* mutant. Three biological replicates of each strain were grown in 50 ml of LB medium or in the defined M9 minimal salts medium supplemented with casamino acids and 0.2% glucose. The OD_{600} was recorded hourly. Growth of a Δ*yhcB* mutant was defective in LB medium, but this defect appeared to be rescued in supplemented minimal medium (Fig. 6.9).

To determine whether the growth defect observed in rich medium was dependent on growth rate, the growth experiment in LB was repeated at 2 additional temperatures, 25°C and 42°C. However, growth of the *yhcB* mutant at 25°C did not rescue the growth defect observed in LB. At both temperatures growth of the Δ*yhcB* mutant was defective relative to the parent strain in LB (Fig. 6.10). This result confirms that the growth defect in rich medium is not due to the increased growth

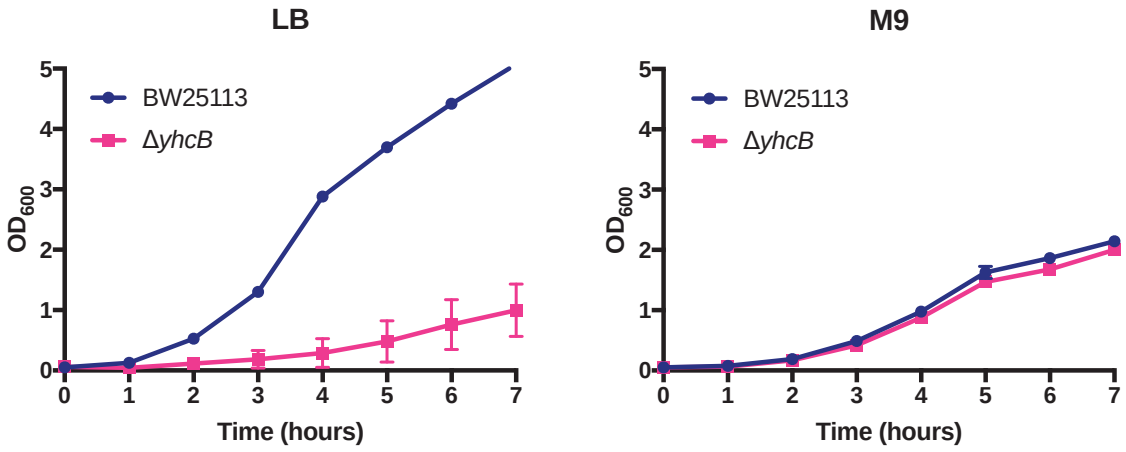

Figure 6.9 Growth kinetics of a Δ*yhcB* mutant

Three replicates of each strain: WT (blue) and Δ*yhcB* (pink), were grown in 50 ml of media in 250 ml flasks and the optical density was recorded hourly for 7 h. The mean and standard deviation of biological replicates were plotted. (A) Growth of strains in LB medium. (B) Growth of strains in M9 medium supplemented with casamino acids and 0.2% glucose. These data were collected by undergraduate student Gavin Cheung.

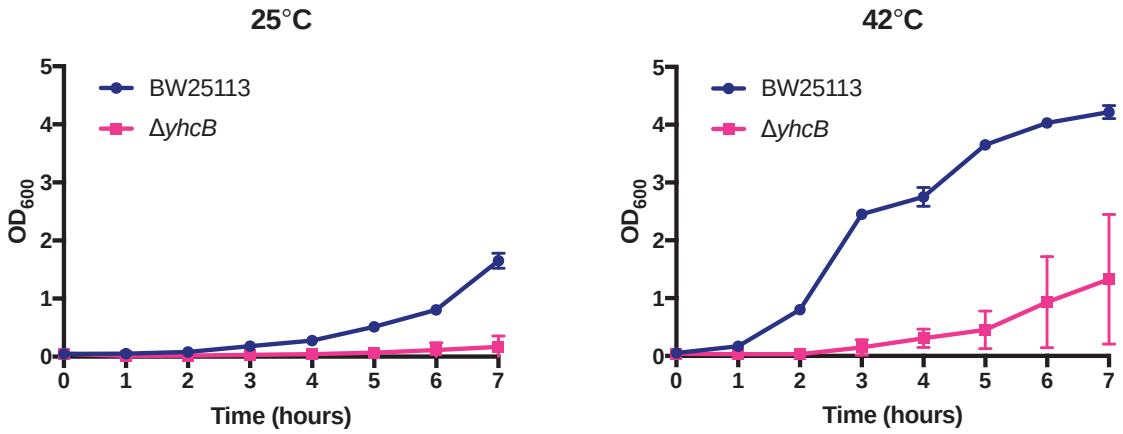

Figure 6.10 An altered growth rate does not restore the growth defect observed in LB

Three replicates of the WT (blue) and Δ*yhcB* mutant (pink) strains were grown in 50 ml LB medium in 250 ml flasks and the optical density was recorded hourly for 7 h. The mean and standard deviation of biological replicates were plotted. (A) Growth of strains at 25°C. (B) Growth of strains at 42°C. These data were collected by undergraduate student Gavin Cheung.

rate, and therefore, might be due to differences in the nutrient components of the media.

6.2.5 The effect of growth medium on the cell morphology of a ΔyhcB mutant

To understand the effect of different media on the growth of a Δ*yhcB* mutant, cells were grown overnight in LB or supplemented M9 minimal salts medium, diluted and visualised using a microscope. The growth defect observed in LB was the same as previous attempts: Δ*yhcB* mutants were visibly larger and irregular in length (Fig. 6.11). However, when grown in M9 minimal salts medium supplemented with glucose, the cell size defect was less apparent (Fig. 6.11). When grown in M9 medium supplemented with glucose, there were still irregular sized cells, but these were less abundant within the total population. The differences in cell morphology when grown in M9 medium plus glucose would explain the difference in growth kinetics observed between growth in the two media.

6.2.6 The effect of carbon source on the growth kinetics of a ΔyhcB mutant

Having established that cell morphology and growth kinetics of a Δ*yhcB* mutant are dependent on media nutrients and not growth rate, the next step was to determine whether the choice of carbon source used to supplement the minimal medium had an effect. Therefore, 4 biological replicates were used to inoculate 200 µl of M9 medium supplemented with 0.2% (v/v) carbon source and incubated in a 96-well plate for 18 hours (the full protocol is described in materials and methods). Under all four conditions, growth of the Δ*yhcB* mutant was defective relative to the WT. However, this was much less severe when the medium was supplemented with glucose compared with the other carbon sources tested (Fig. 6.12). These data suggest that

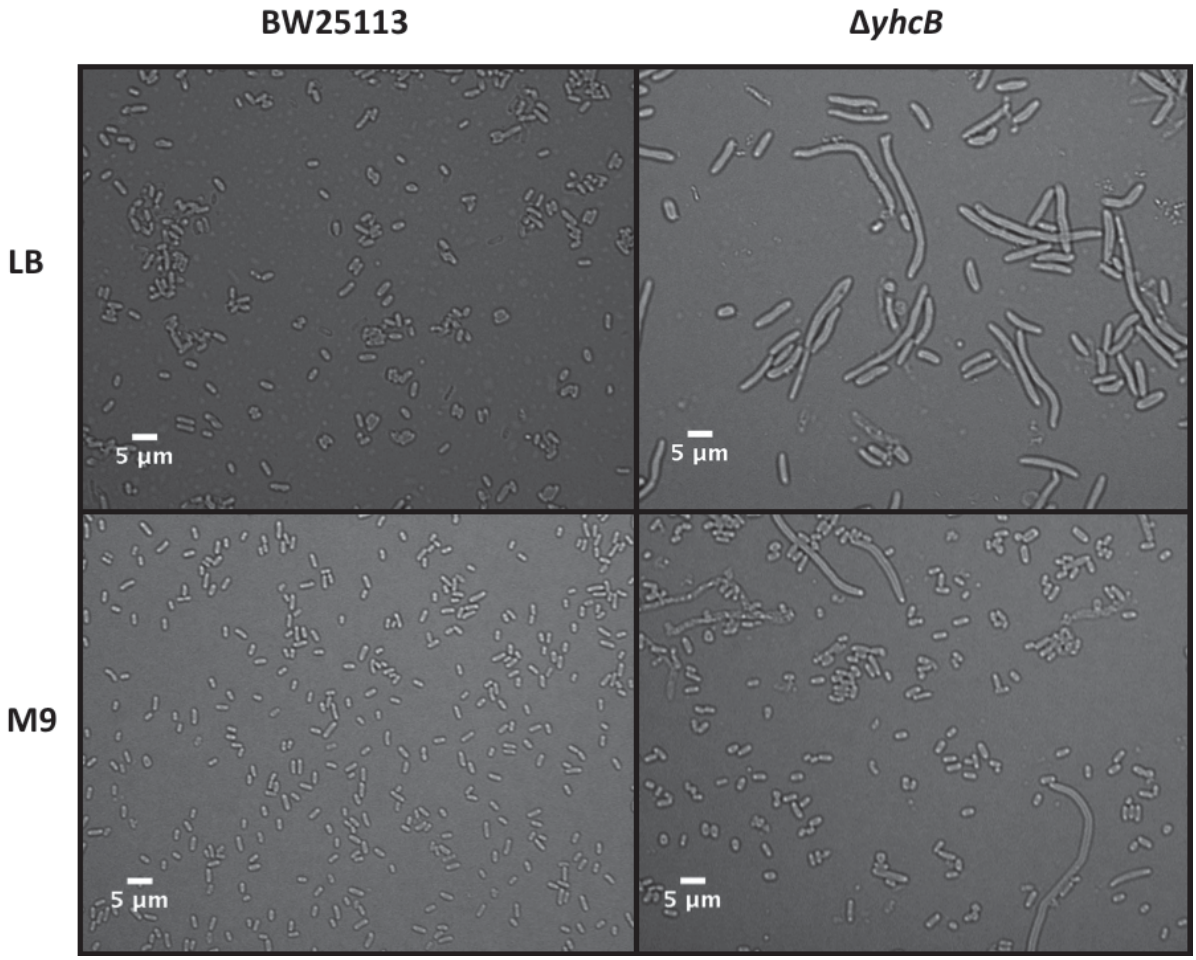

Figure 6.11 The effect of media on cell morphology

Overnight cultures of strains were grown in LB or M9 minimal medium supplemented with casamino acid and 0.2% glucose, and diluted for visualisation. Cells were viewed using a Nikon 90i eclipse microscope. A 5 µm size bar is shown for reference, the DIC images presented were captured using a 40x objective lens.

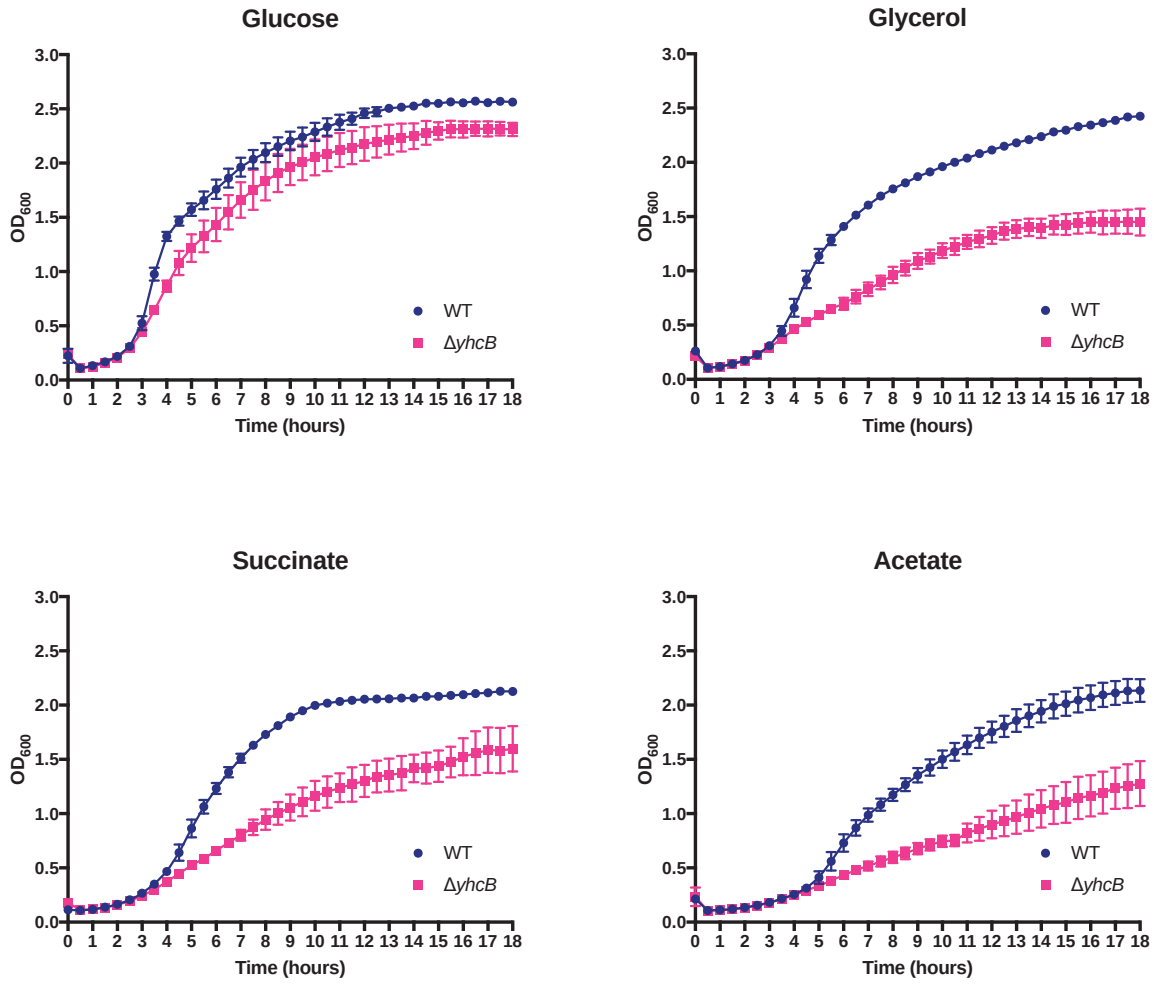

Figure 6.12 The effect of carbon source on the growth of a Δ*yhcB* mutant

Four biological replicates were grown in M9 minimal salt medium supplemented with casamino acids and 0.2% carbon source. Strains were grown in 200 µl media in a 96-well plate and the OD_{600} was recorded at 30 min intervals for 18 h. The mean and standard deviation of biological replicates were plotted; WT (blue) and Δ*yhcB* (pink).

the semi-restored phenotype of a *yhcB* mutant grown in supplemented M9 medium compared to LB (Fig. 6.9 & 6.11) is due to the supplemented glucose and not additional components within the M9.

6.2.7 Complementation of a ΔyhcB mutant

To complement the effect of deleting *yhcB*, the *yhcB* gene was cloned into the pBAD vector downstream of an arabinose inducible promoter and upstream of, and in frame with, a Myc and 6x poly-His tags (Fig. 6.13A). Two variants of the plasmid were constructed, either with or without a stop codon, resulting in a Myc-His tagged and a non-tagged variant, respectively. To confirm expression of the YhcB protein from these plasmids and correct fusion of the tag, cells carrying either an empty pBAD plasmid (empty vector: E), pBAD::yhcB-Myc-His (H) or pBAD::yhcB (X) were grown overnight with glucose and then sub-cultured into fresh medium supplemented with 0.4% arabinose to induce expression from the P_{BAD} promoter. After two hours of growth, whole cells harvested, lysed and the total cellular proteins separated by SDS-PAGE and visualised by Coomassie blue staining. There was an abundant small protein in each of the lanes originating from the strains carrying plasmids with cloned copies of *yhcB*. The two bold bands were therefore assumed to be YhcB (Fig. 6.13B). The overabundant protein in lane X migrated further down the gel than the most abundant protein in lane H, which corresponds to the size difference expected for the addition of two tags. Western blot analysis with an anti-His antibody confirmed the presence of a His tag, and this was only detected in lane H. An additional, fainter His signal was detected further up the lane for a protein around twice the size of YhcB (Fig. 6.13C). Despite using denaturing conditions to separate proteins during PAGE

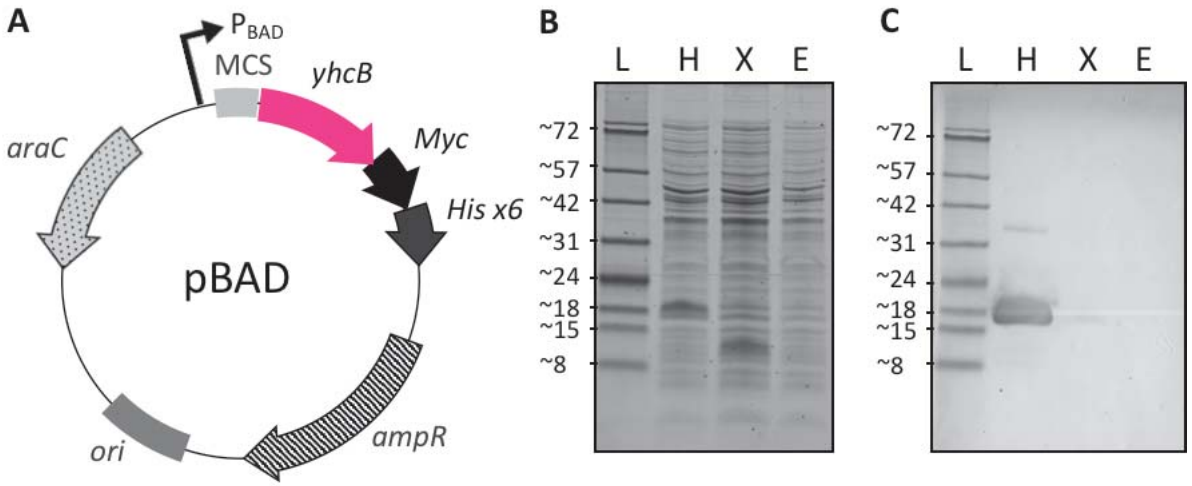

Figure 6.13 Cloning *yhcB* into the pBAD vector.

(A) *yhcB* was cloned into the pBAD vector, in frame with the Myc tag and poly-His tag. Two vectors were constructed: the *yhcB* gene was cloned with and without its stop codon (H and X respectively). (B) Coomassie-blue staining of the proteins extracted by whole cell lysate from BW25113 carrying pBAD vectors H, X or E (empty vector). The proteins were separated in a 15% polyacrylamide gel. (C) Anti-His western blot analysis of the same whole cell lysates separated by SDS-PAGE and transferred to a membrane.

analysis, the fainter, larger band is thought to be a YhcB dimer, given the strong self-interacting BACTH data reported by Li *et al.* (2012).

Having confirmed that the expression plasmids produce YhcB when induced, the aim was to use these plasmids to induce YhcB expression to complement Δ*yhcB* sensitivity to SDS + EDTA. The plasmids were transformed into both WT cells and Δ*yhcB* cells. After overnight growth, transformants were spotted onto solid medium supplemented with glucose or a range of arabinose concentrations (Fig. 6.14). YhcB provided on a plasmid was able to restore the SDS + EDTA sensitivity phenotype, as cells were no longer ~1,000-fold more sensitive to SDS + EDTA, and had a comparable sensitivity profile to WT cells with the empty vector control. However, irrespective of whether YhcB expression was induced in WT or Δ*yhcB* cells, the overall population morphology was altered as the culture spots appeared translucent (Fig. 6.14). This suggested that the additional copies of YhcB were able to suppress the SDS + EDTA sensitivity phenotype, but resulted in an altered cell morphology compared to unstressed WT cells. The effect was most pronounced at higher concentrations of arabinose (Fig. 6.14). Growth of WT cells in liquid broth, when expressing YhcB from the complementation plasmid, was retarded relative to the WT cells carrying empty pBAD plasmid (Fig. 6.15). This suggests that overexpression of YhcB rescues specific defects in a Δ*yhcB* background in response to SDS + EDTA but additional copies of YhcB in a WT background have a fitness cost.

To support this, WT and Δ*yhcB* cells carrying empty pBAD vector or pBAD-yhcB* were visualised using a microscope 2 hours after induction with 0.4% arabinose. The phenotypes of WT and Δ*yhcB* cells carrying an empty pBAD vector were the same as previously reported. However, the size of Δ*yhcB* cells complemented with pBAD-yhcB* appeared less defective, whereas WT cells complemented with pBAD-yhcB*

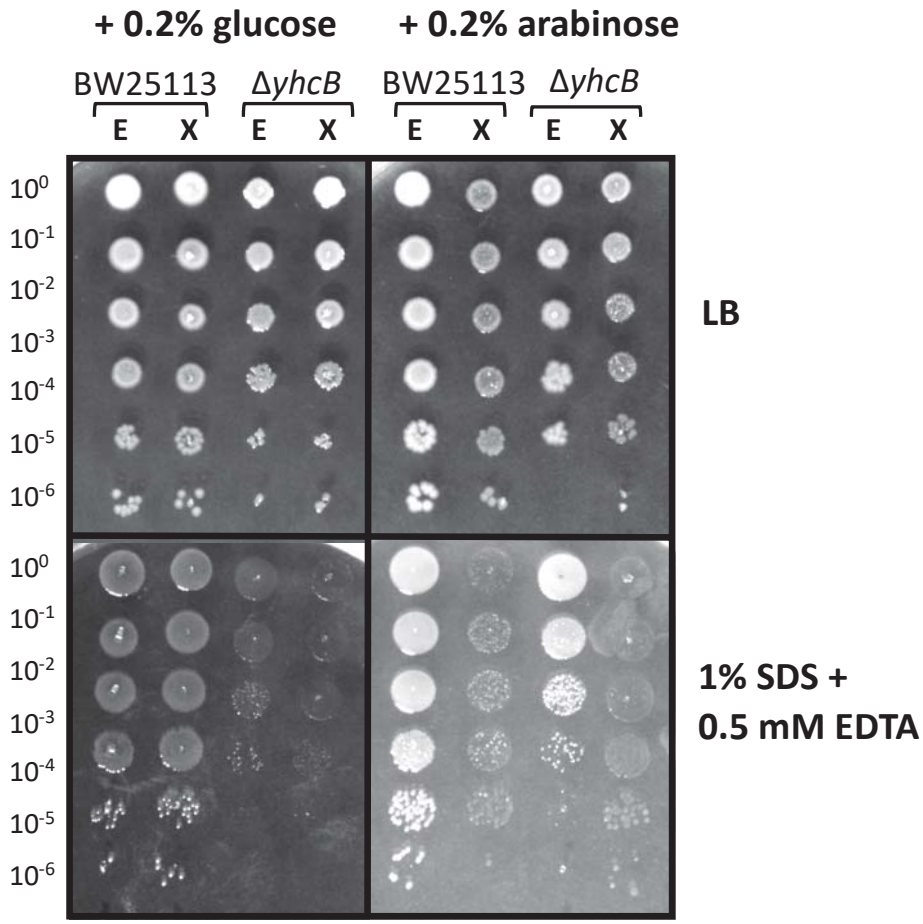

Figure 6.14 Overexpression of YhcB affects colony morphology and strain sensitivity to stress

WT and Δ*yhcB* strains transformed with empty pBAD vector (E) or pBAD::yhcB (X) were 10-fold serially diluted from an initial OD_{600} of 1.00. 2 µl of cells were spotted onto LB agar plates supplemented with 0.2% glucose, or 0.2% arabinose, with and without 1% SDS + 0.5 mM EDTA.

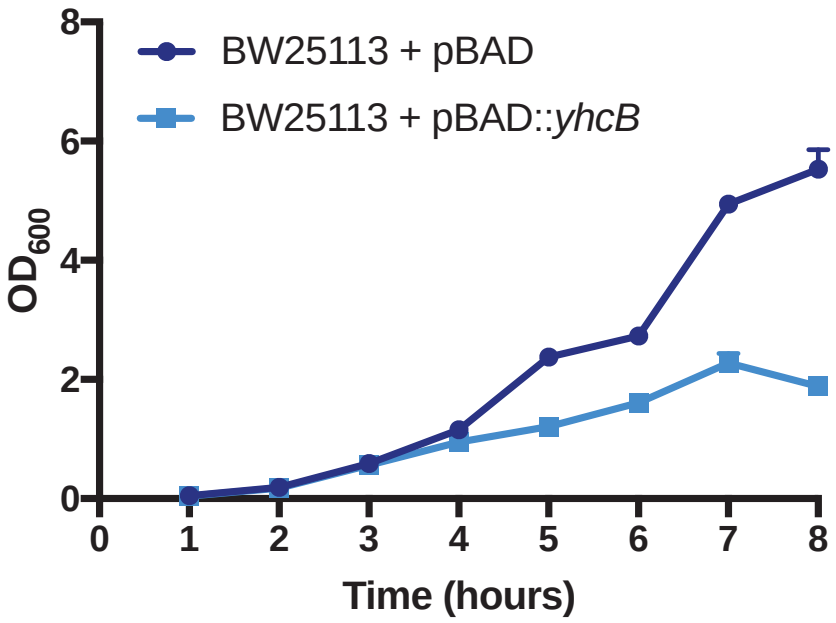

Figure 6.15 The growth kinetics of a WT strain overexpressing YhcB

Three replicates of each strain, WT + pBAD (dark blue circles), WT + pBAD::*yhcB* (light blue squares), were grown in 50 ml LB supplemented with 0.2% arabinose in 250 ml flasks and the optical density was recorded hourly for 8 h. Overexpression of *yhcB* in a WT background retards growth. The mean and standard deviation of biological replicates were plotted. These data were collected by undergraduate student Gavin Cheung.

appeared to have a defective cell length and an increased proportion of longer cells, suggesting a mildly defect in division timing (Fig. 6.16). Altogether, these data indicate that the effects of *yhcB* deletion can be complemented, but the total number of YhcB proteins within a cell is important for correct cell size and division.

6.2.8 Identification of synthetic lethal partners of yhcB

In addition to BACTH screens, Li *et al.* (2012) also reported that RodZ is a synthetically lethal partner of YhcB. TraDIS has previously been used as a tool to identify synthetic lethal partners on a genome scale (DeJesus *et al.* 2017). To identify additional genes with a synthetically lethal relationship with *yhcB*, a transposon mutant library was constructed in a Δ*yhcB* background, using a mini Tn*5* transposon with a kanamycin resistance cassette. Approximately 800,000 mutants were pooled and sequenced, resulting in identification of 360,840 unique insertion sites from 4,837,705 mapped reads. This corresponds to an average of one insertion approximately every 13 bp. The insertion sites were evenly distributed around the genome (Fig. 6.17).

No insertions were found within *yhcB*, which was a good indication that the library had been constructed successfully. The number of insertions per CDS were quantified (using the same method as previously) and 409 genes were identified as essential (S. Table 6.1). The essential genes identified in the Δ*yhcB* library were compared to essential genes of a BW25113::Tn*5*-kanamycin library as a control (Fig. 6.18). There was an overlap of 305 essential genes between the two datasets. There were an additional 104 genes that are conditionally essential in a Δ*yhcB* background, and 23 genes predicted to be essential in a WT library but non-essential when *yhcB* is deleted (S. Table 6.2). The large

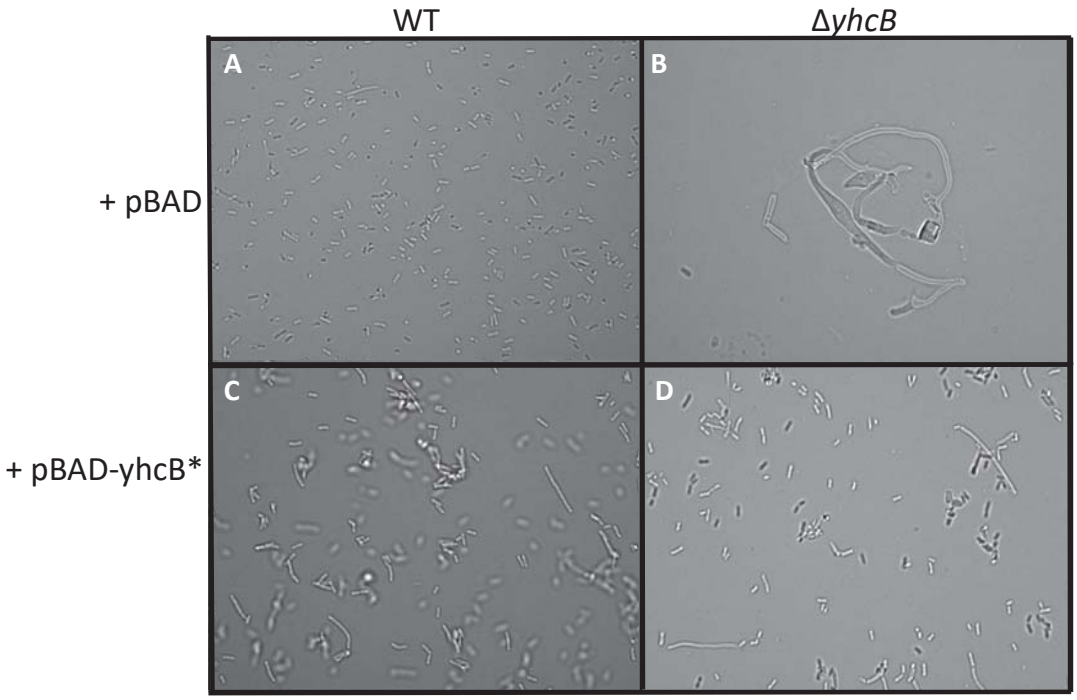

Figure 6.16 Complementation of the cell morphology defect in a Δ*yhcB* mutant

WT and Δ*yhcB* cells transformed with empty pBAD vector or pBAD::*yhcB* were induced with 0.4% arabinose and viewed using a Nikon 90i eclipse microscope after 2 h growth. The DIC images presented were viewed through the 40x optical lens.

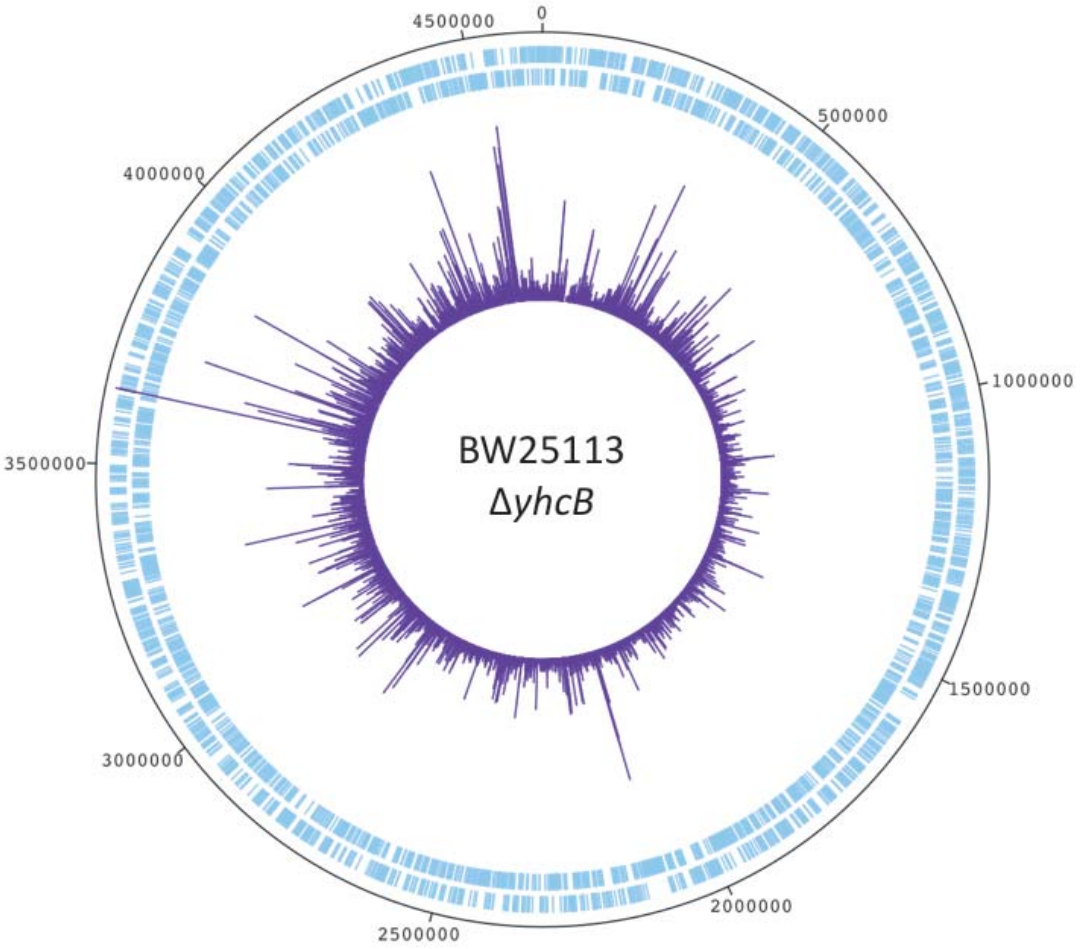

Figure 6.17 Construction of a transposon mutant library in a Δ*yhcB* strain

A genome map of BW25113 starting at the annotation origin. The sense and antisense coding sequences of BW25113 are shown in blue, respectively. The position and frequency of sequenced transposon insertion events are shown in purple.

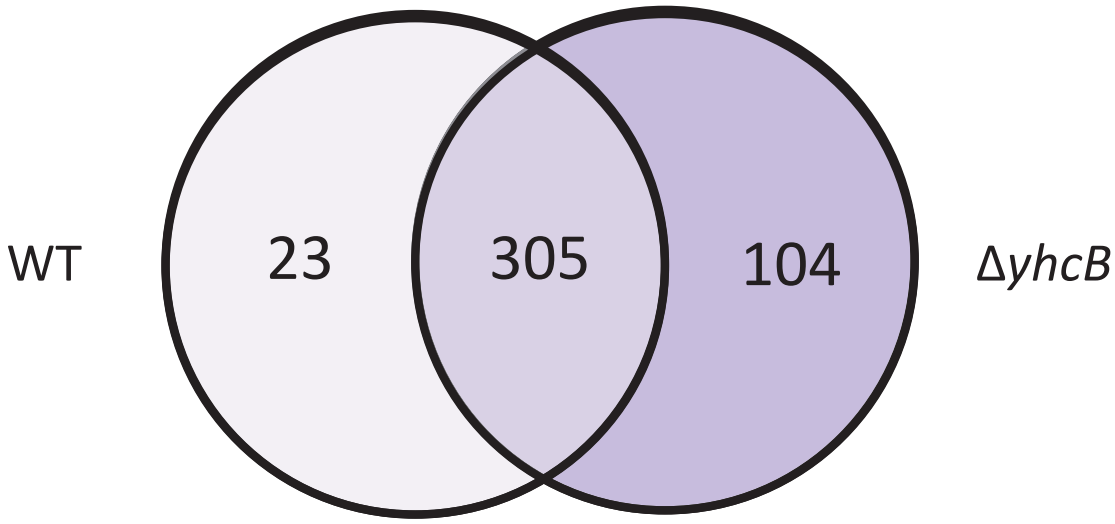

Figure 6.18 Comparison of essential genes between a BW25113 and a BW25113Δ*yhcB* transposon library

Essential genes were compared between a transposon library constructed in BW25113 (light purple) and a transposon library constructed in a Δ*yhcB* background (purple). There was an overlap of 305 essential genes in both libraries and an additional 104 essential genes in the Δ*yhcB* library (dark purple). There are 23 genes that were no longer essential in a Δ*yhcB* background (light purple).

number of genes that are synthetically lethal in combination with *yhcB* is not surprising given that the defective morphology of a *yhcB* mutant.

Because *rodZ* had previously been identified as a synthetic lethal partner of *yhcB*, the list of synthetically lethal genes were checked for the presence of *rodZ* to validate the library construction, however, *rodZ* was not initially identified. On closer inspection, only the 5′ end of the gene is conditionally essential, corresponding with the cytoplasmic and transmembrane domains of the RodZ protein (Fig. 6.19). Insertions within the latter half of the gene result in an IIS that does not meet criteria for *rodZ* to be classified as essential by the analysis method used. The TraDIS data confirms that *rodZ* and *yhcB* are synthetically lethal, but specifically the transmembrane and cytoplasmic domains of rodZ are essential in a *yhcB* mutant.

6.2.9 Functional enrichment of synthetically lethal partners of yhcB

To determine whether there were any common pathways or functions that were enriched among the conditionally essential genes in a Δ*yhcB* background, the 104 genes were processed using the online Gene Ontology (GO) database (Ashburner *et al.* 2000; The Gene Ontology Consortium 2017). Every gene within *E. coli* is assigned one or more corresponding GO categories based on the function of the protein that the gene encodes. Therefore, the proportion of genes that code for products in a certain pathway or cellular process, relative to the whole genome, is quantifiable. The PANTHER overrepresentation test was used to compare the expected proportion of genes in a given category to the actual number of identified genes per GO category, which can be used to calculate a fold-enrichment score with an associated p-value (Mi *et al.* 2017).

Figure 6.19 The dispensable and essential domains of *rodZ* within a Δ*yhcB* mutant

A screenshot of the artemis genome browser showing the location and frequency of identified transposon insertion sites within the Δ*yhcB* transposon library (pink). The *rodZ* gene is shown in light blue. There is a clear window at the 5′ end of the *rodZ* gene were there are no transposon insertion sites. This corresponds with the portion of the *rodZ* CDS that codes for the cytoplasmic and transmembrane (TM) domains (purple and dark blue, respectively), and a fraction of the periplasmic domain of RodZ.

The functions of the 104 conditional essential genes of a Δ*yhcB* strain were compared to the total cell protein function categories of a whole cell. A total of 36 GO categories, with a false discovery p-value of less than 0.05, were identified as significantly enriched within the Δ*yhcB* data and are presented in Appendix 6.1. A number of these categories are functionally related and include the same genes, or combinations of similar genes, so the GO group with the highest fold-enrichment score per overarching category was isolated. These are presented in figure 6.20. These data suggest that the processes 'protein transport', 'bacteriocin transport', 'ECA (Enterobacterial common antigen) biosynthesis', 'LPS core biosynthesis' and 'cell division' are functionally important in a Δ*yhcB* cell (Fig. 6.20). The data, however, are biased for groups of a smaller size. For example, the groups 'protein import' and 'bacteriocin transport' have a total of 4 and 6 proteins each, and 3 (TolABQ) and 4 (TolABQR) of these proteins, respectively, were in the list of Δ*yhcB* synthetic lethal partners. Furthermore, although the GO functions are separate, the contents are overlapping and results in the reporting of some proteins twice. Similarly, the number of proteins in the cell division category is 54, of which 8 were identified in the Δ*yhcB* data (TolA, TolB, TolQ, TolR, Pal, MepM, MatP and DacA). Given that the Tol-Pal system is required for correct invagination of the OM during division and a Δ*yhcB* mutant has a defective OM but does not localise to the septum, it is likely that *yhcB* is not functionally related to the Tol-Pal system and the synthetic lethality is due to defective cell division (Gerding *et al.* 2007). The enrichment of genes involved in the synthesis of enterobacterial common antigen (ECA) and LPS-core were investigated further.

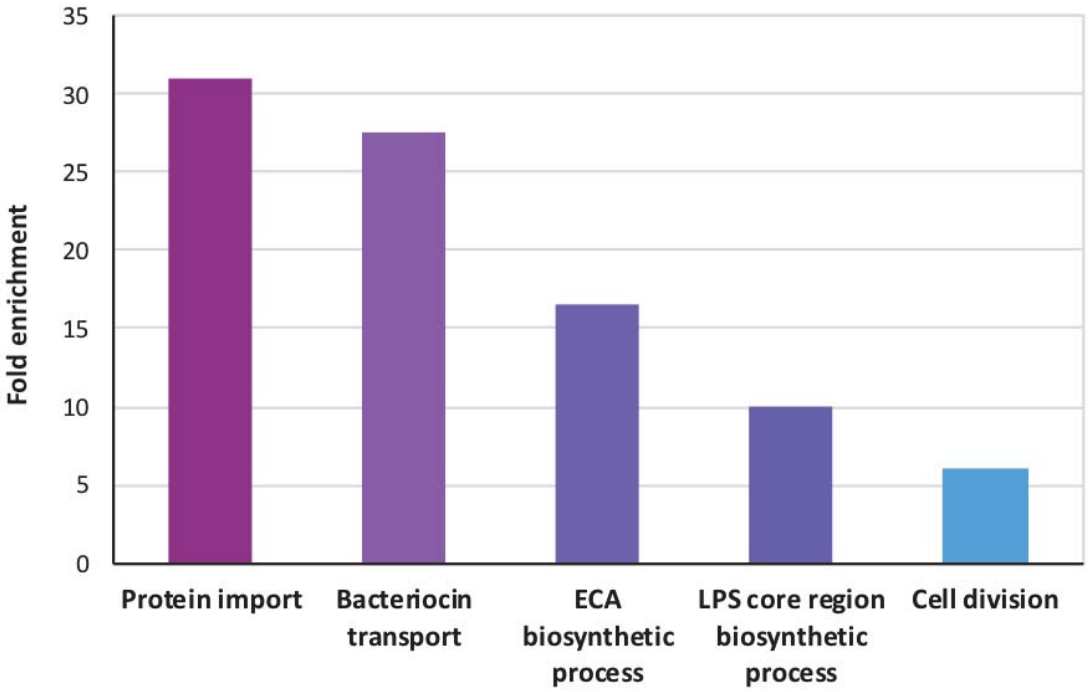

Figure 6.20 Functional enrichment of genes that are synthetically lethal with *yhcB*

The fold enrichment of significantly overrepresented gene ontology (GO) groups within the list of genes that are conditionally essential in a Δ*yhcB* strain. GO groups are shown along the x-axis.

6.2.10 Characterisation of the conditionally essential LPS biosynthesis genes

Following the identification of the enrichment of LPS-core biosynthesis genes among the synthetic lethal partners of *yhcB*, the overall LPS biosynthetic pathway was inspected further (Fig. 6.21A). A schematic of LPS structure is shown in figure 6.21B). Enzymes involved in LPS biosynthesis are shown next to the link they form (central column) or the component they ligate (side branches; Fig. 6.21B). Essential enzymes are indicated in black, conditionally essential enzymes are in pink. The first non-essential step in LPS biosynthesis, the acylation of lipid A by LpxM, is conditionally essential in a Δ*yhcB* strain (Fig. 6.21C). All heptose components of the inner core and the entire heptose biosynthetic pathway are essential (Fig. 6.21D). However, the third Kdo moiety is not essential and neither is the outermost phosphate ligated by WaaY. Finally, the last step in the synthesis of D-glucose is essential, as is the enzyme WaaG, which ligates the first glucose of the outer core. Together these data suggest that the first glucose of the outer core is essential, but the remainder of the outer core is not. These data also suggest that the inner core imparts additional structural integrity that is especially important in a Δ*yhcB* strain.

6.2.11 Characterisation of the conditionally essential genes of the ECA pathway

The ECA biosynthetic pathway was also inspected further to understand why genes involved in ECA biosynthesis might be conditionally essential in a Δ*yhcB* mutant. A schematic of the ECA biosynthetic pathway was adapted from Jorgenson *et al.* (2016). The conditionally essential steps are highlighted in pink (Fig. 6.22). TraDIS data from the Δ*yhcB* library suggest that all steps after the synthesis of ECA-lipid I are lethal in a Δ*yhcB* strain. However, disruption of *wzzE* results in viable mutants. WzzE mediates

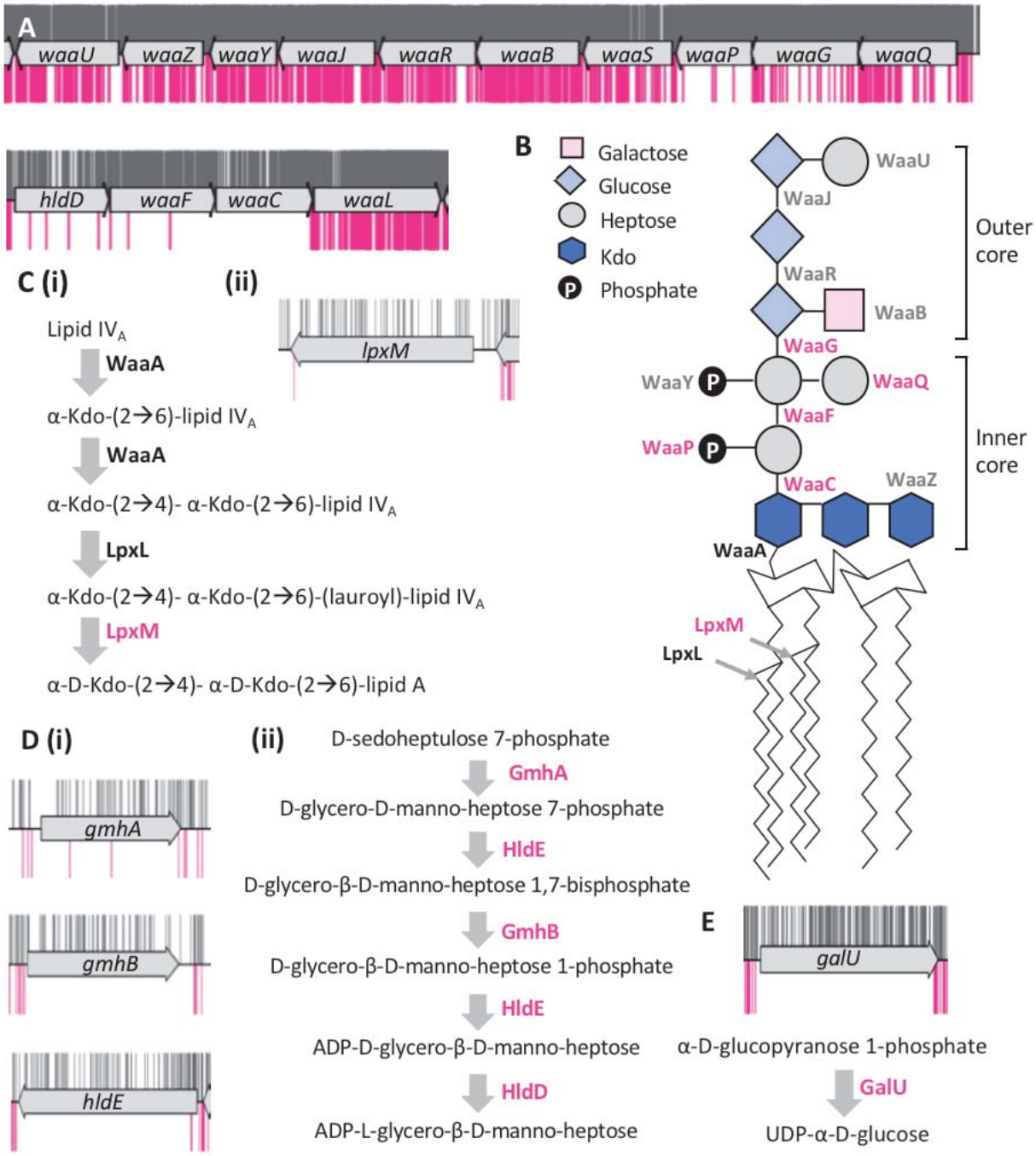

Figure 6.21 Characterisation of the LPS core biosynthesis genes that are conditionally essential in a Δ*yhcB* strain

TraDIS data for BW25113::Tn-kan (grey) and BW25113Δ*yhcB*::Tn-kan (pink) were plotted above and below the gene track, respectively. Insertion sites are represented by vertical bars and are capped at a frequency of 1. (A) Insertions within genes required for LPS core biosynthesis are significantly underrepresented.

(B) A schematic of the structure of LPS. LPS biosynthesis enzymes are indicated next to the linkage they form (central column) or component they ligate (side branches). Conditionally essential enzymes are labelled in pink, essential enzymes in black and non-essential enzymes in grey. (C) The final step in the synthesis of mature lipid A is essential in a Δ*yhcB* background. (D) The committed steps of the heptose biosynthesis pathway are conditionally essential. (E) The final step in the synthesis of D-glucose is conditionally essential.

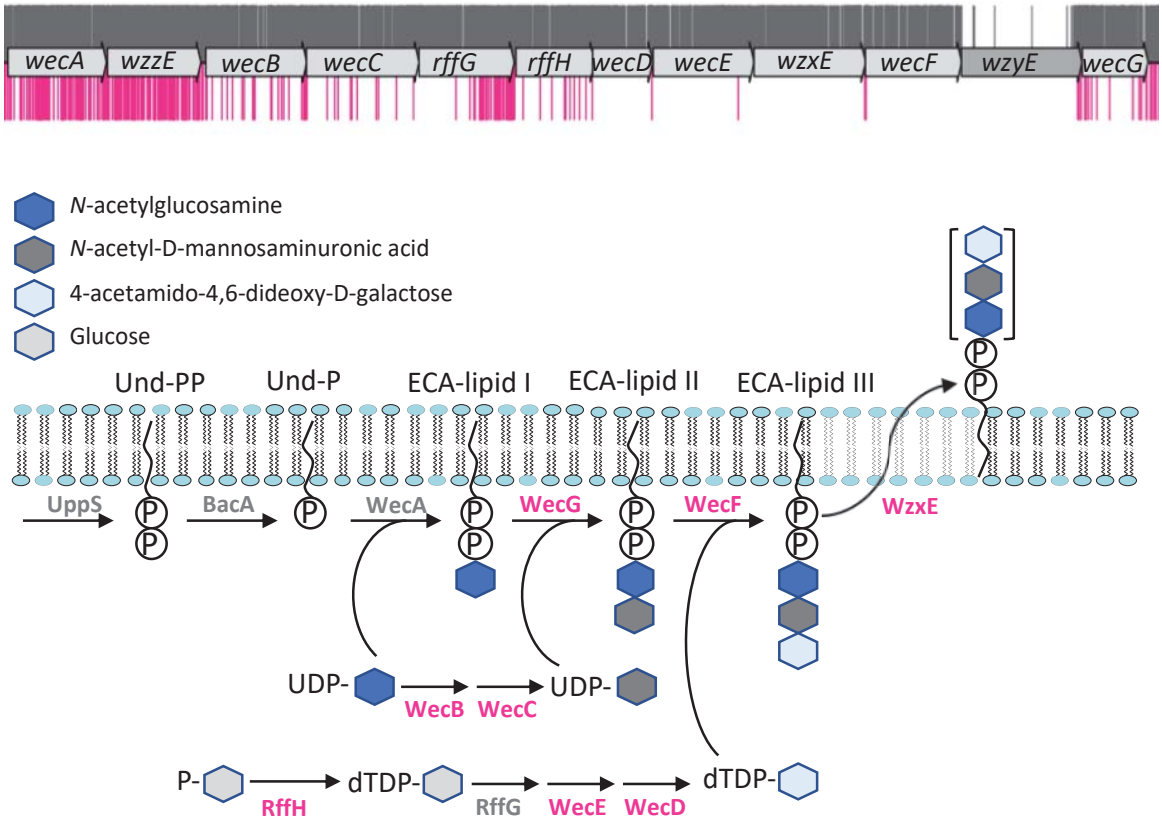

Figure 6.22 Characterisation of the synthetically lethal genes involved in ECA biosynthesis

(A) TraDIS data for the enterobacterial common antigen (ECA) biosynthesis operon. BW25113::Tn-kan (grey) and BW25113Δ*yhcB*::Tn-kan (pink) data were plotted above and below the gene track, respectively. Insertion sites are represented by vertical bars and are capped at a frequency of 1. Core-essential genes are shown in darker grey. (B) A schematic of the ECA biosynthesis pathway adapted from Jorgenson *et al.* (2016). Conditionally essential enzymes are shown in pink, non-essential enzymes are shown in grey. Abbreviations: uridine diphosphate (UDP), thymidine diphosphate (dTDP), undecaprenyl phosphate (Und-P), undecaprenyl pyrophosphate (Und-PP)

the length of the ECA chains by determining the number of repeating units (Barr *et al*. 1999). Similarly, disruption of *wecA* results in viable mutants. WecA catalyses the first committed step in ECA biosynthesis, which is the transfer of *N*-acetylglucosamine-1-phosphate onto undecaprenyl phosphate (Und-P) to form ECA-lipid I (Fig. 6.22). In short, an ECA-null mutant is viable and mutants with variable ECA lengths are also viable; only mid-pathway blocks are lethal. A Δ*wecE* mutant is known to have a longer, swollen cell morphology (Jorgenson *et al*. 2016). The phenotype of a Δ*wecE* mutant is due to accumulation of undecaprenyl linked ECA-lipid II intermediates (Jorgenson *et al*. 2016). By extension, the TraDIS data reveal the steps of the ECA biosynthetic pathway that, when disrupted, result in lethal accumulation of ECA-lipid intermediates.

Either sequestration of Und-P (a compound of limited abundance that sits at the start of several cell envelope biosynthetic pathways) in a dead-end pathway is lethal, suggesting that the availability of Und-P might be more limited in a Δ*yhcB* mutant. Or, the OM perturbations that result from disruption to the ECM biosynthetic pathway have a cumulative effect alongside the OM defects of a *yhcB* mutant, which is lethal (Danese *et al*. 1998; Ramos-Morales *et al*. 2003; Jorgenson *et al*. 2016).

6.2.12 Characterisation of peptide crosslinking mutants within a ΔyhcB mutant

The overall shape of a cell is often governed by the cell wall. As a *yhcB* mutant has such an aberrant cell shape, occasionally with visible bulges along the length, and cells are prone to lysis, it is therefore possible that the root of the Δ*yhcB* cell envelope defect is due to a defect in cell wall biosynthesis. Most genes involved in cell wall biosynthesis are core essential genes, and so there are limited candidates for conditionally-essential genes. However, the list of conditionally essential genes in a

Δ*yhcB* mutant was inspected for genes involved in peptidoglycan biosynthesis. Four genes (*mepM*, *mepS*, *dacA* and *dapF*) with roles in cell wall biosynthesis were identified within this list (Fig. 6.23). The genes *mepM* and *mepS* code for 2 of the 3 peptidoglycan endopeptidases in *E. coli* that cleave the 3-4 *meso*-Dap-D-Ala crosslink. The 3rd gene, *mepH,* does not appear essential. Together, the three *mep* genes are synthetically lethal, but over-expression of any one of them will rescue this phenotype indicating functional redundancy (Singh *et al*. 2012).

An additional carboxypeptidase (LdcA) involved in murein recycling in the cytoplasm is also conditionally essential (Data not shown). Deletion of LdcA in WT cells results in a cell wall crosslinking defect. The data were initially counterintuitive, it might be expected that in a cell with decreased envelope integrity, the loss of endopeptidase activity might be beneficial as glycan strands would remain structurally crosslinked. However, from these data, it could be inferred that cell wall synthesis is impeded in a Δ*yhcB* mutant, and further defects in PGN turnover are lethal.

6.2.13 Transposon insertion mutations that restore the sensitivity of ΔyhcB mutants to OM stress

To identify insertion mutants that are resistant to vancomycin or SDS and EDTA, and therefore suppress a Δ*yhcB* phenotype, the mutant library was screened on LB agar supplemented with varying compound concentrations (Fig. 6.24). The conditions used were: 1% SDS + 0.5 mM EDTA; 1% SDS + 1 mM EDTA; 200 µg/ml vancomycin and 300 µg/ml vancomycin. All plates were inoculated with approximately the same number of cells. Cultures were diluted to comparable cell numbers and spotted onto plates in such quantities that single colonies of transposon mutants could be isolated,

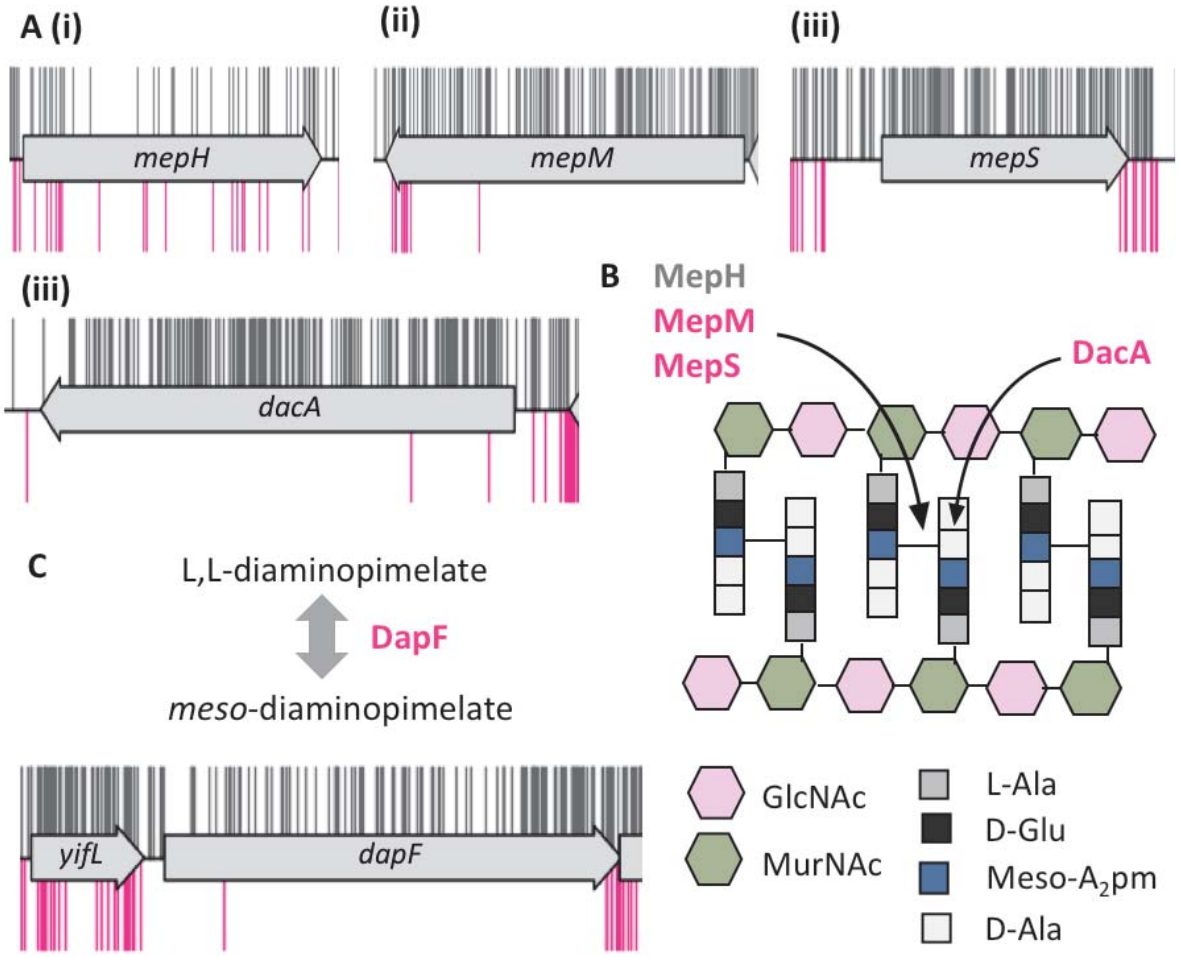

Figure 6.23 Characterisation of peptide crosslinking mutants within a Δ*yhcB* strain

TraDIS data of the BW25113::Tn-kan library (grey) and BW25113Δ*yhcB*::Tn-kan library (pink) were plotted above and below the gene track, respectively. Insertion sites are represented by vertical bars and are capped at a frequency of 1. (A) TraDIS data for the following: (i-iii) Peptidoglycan endopeptidases that cleave D-ala-*meso*-Dap cross links (iv) DD-carboxydpeptidase, DacA, which cleaves D-Ala-D-Ala linkages releasing the terminal D-Ala residue. (B) Schematic representation of the target sites of the endopeptidases and carboxypeptidase. (C) TraDIS data for DapF, an epimerase that synthesises the conversion of L,L-diaminopimelate to *meso*-diaminopimelate (Meso-A_2pm).

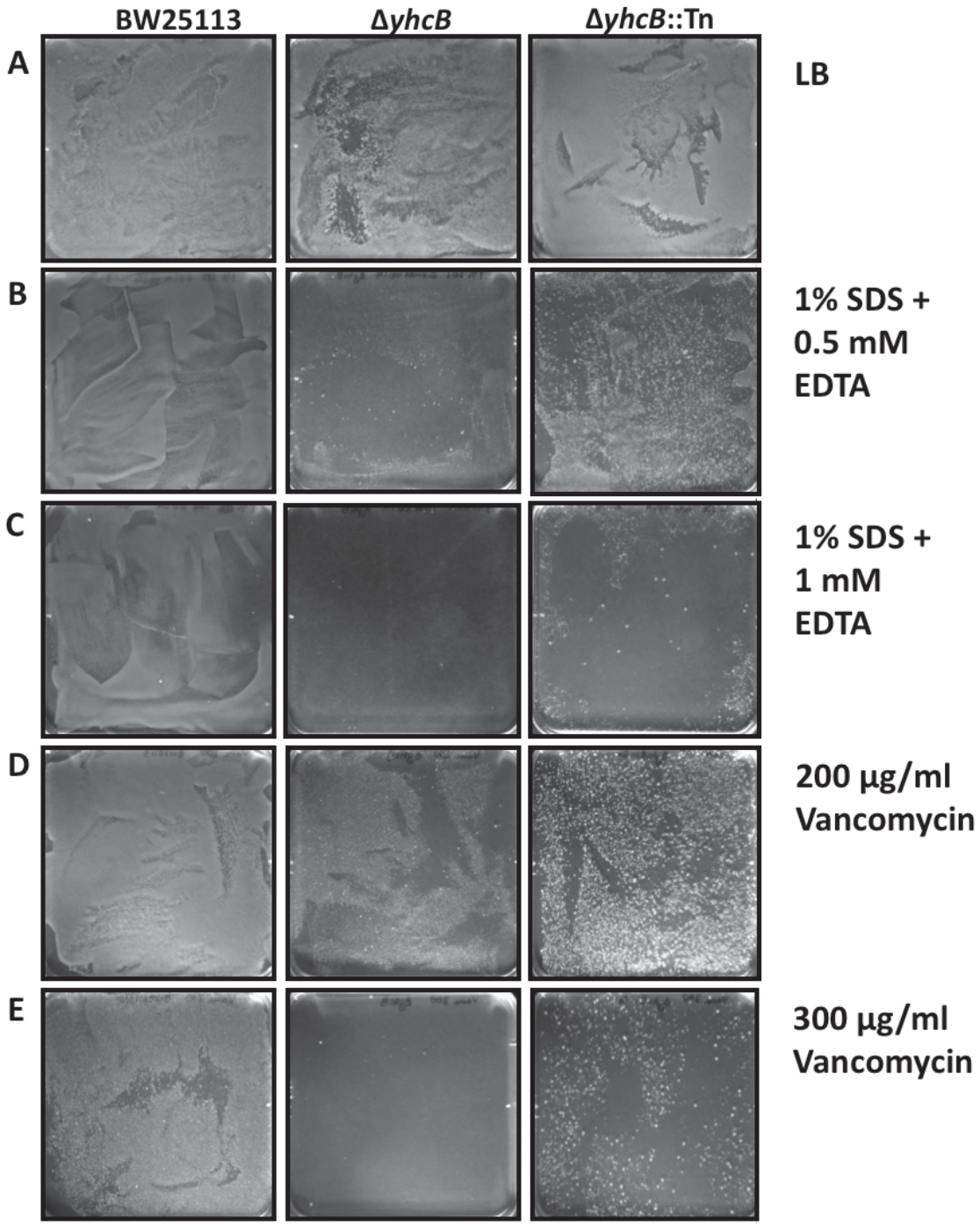

Figure 6.24 Isolation of Δ*yhcB*::Tn suppressor mutants

Comparable numbers of cells were inoculated onto plates of LB (A), or supplemented each with 2 different concentrations of SDS + EDTA (B & C), or vancomycin (D & E).

to prevent the growth of satellite colonies and therefore false positive mutants. Under these conditions the WT strain formed a lawn while growth of the Δ*yhcB* strain was limited (Fig. 6.24). The Δ*yhcB* cells formed a lawn only on LB, while growth on supplemented plates was either sparse and limited to a small colony phenotype or non-existent, except for a few isolated large colonies presumed to be natural suppressor mutants (Fig. 6.24D and E). Meanwhile, under these conditions there was visibly more growth of transposon mutants relative to the parent Δ*yhcB* strain suggesting that there were transposon mutants within the library able to suppress the OM defects of the parent strain. These mutants were collected and sequenced to identify the transposon insertion sites.

Insertion sites with fewer than 10 mapped reads were disregarded to filter for noise. Genes with 10 or more mapped reads per insertion site along the full length of the gene were recorded. These genes are considered to be gene-deletion suppressor mutants. There were 49 genes that, when disrupted, conferred resistance to 1% SDS + 0.5 mM EDTA. This number decreased to 27 genes when the concentration of EDTA was increased to 1% SDS + 1 mM EDTA. Similarly, there were 25 genes that conferred resistance to 200 µg/ml vancomycin, which decreased to 21 genes when the concentration of vancomycin was increased to 300 µg/ml. A four-way comparison between these genes revealed only 2 genes that were common to all four datasets, *nlpI* and *rsmA* (Fig. 6.25). The transposon insertion data for these genes are shown in figure 6.26A, only the data for the two lower concentrations of each stress are shown. NlpI is an OM lipoprotein. NlpI mediates the degradation of MepS by the protease Prc. In a Δ*nlpI* strain, MepS activity is significantly increased (Singh *et al.* 2015). A Δ*nlpI* strain also has high levels of peptidoglycan synthesis, and has an increased level

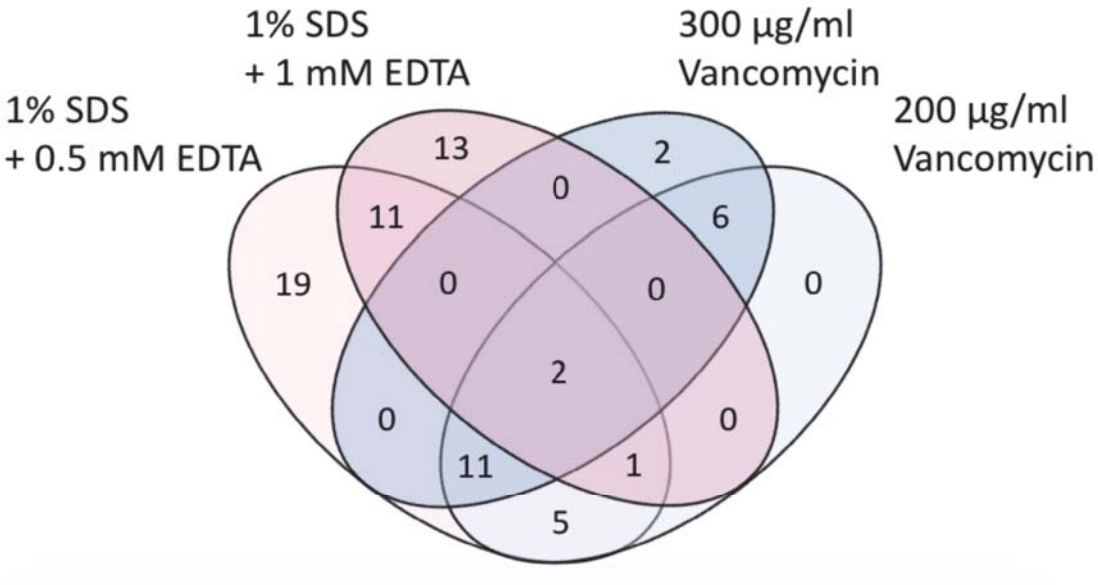

Figure 6.25 Comparison of transposon suppressor mutants between different conditions

Four-way comparison of the gene-disruption suppressor mutants identified after growth of a transposon mutant library on LB agar plates supplemented with increasing concentrations of SDS + EDA or vancomycin. Genes isolated in this analysis are considered gene-deletion suppressors if transposon insertions were identified along the full length of the CDS. To account for noise within the data, insertion sites with fewer than 10 mapped reads were discarded.

of OM vesicles as a consequence (Schwechheimer *et al.* 2015). The gene *rsmA* encodes a methyltransferase that methylates 16S rRNA; it is not clear why deletion of *rsmA* might suppress the defects of a *yhcB* deletion mutant (Poldermans *et al.* 1979).

A surprising result from the TraDIS suppressor data was the identification that disruption of any component of the Mla pathway appears to suppress the sensitivity of a *yhcB* mutant to SDS + EDTA (Fig. 6.26). This was unexpected as *mla* mutants are highly sensitive to SDS + EDTA (Ekiert *et al.* 2017). These data suggest that, while a single *yhcB*-deletion mutant or *mla* mutant might be sensitive to SDS + EDTA, if a double mutant is constructed, resistance is restored.

6.2.14 Spontaneous suppressor mutations that suppress the sensitivity of a ΔyhcB mutant to vancomycin

During screening of the transposon library on vancomycin, 6 revertant suppressors of Δ*yhcB* were isolated that were able to grow on LB supplemented with vancomycin. Two colonies were isolated from the plate supplemented with 200 µg/ml vancomycin, and 4 colonies were isolated from the plate supplemented with 300 µg/ml vancomycin. To confirm resistance to vancomycin, 10-fold serial dilutions of culture were inoculated onto LB, and LB supplemented with vancomycin or SDS and EDTA. As expected, all 6 mutants were equally or more resistant to vancomycin than the parent strain and the WT (Fig. 6.27B). Interestingly, some of these mutants were also able to suppress on SDS + EDTA (Fig. 6.27C). Suppressor mutants number 3 and 5 had the greatest level of suppression under both conditions. To confirm the mutants were indeed Δ*yhcB* mutants, the *yhcB* gene locus was amplified by PCR using primers that flank the *yhcB* gene (Fig. 6.27D). All 6 suppressor mutants were confirmed to be *yhcB* deletion mutants.

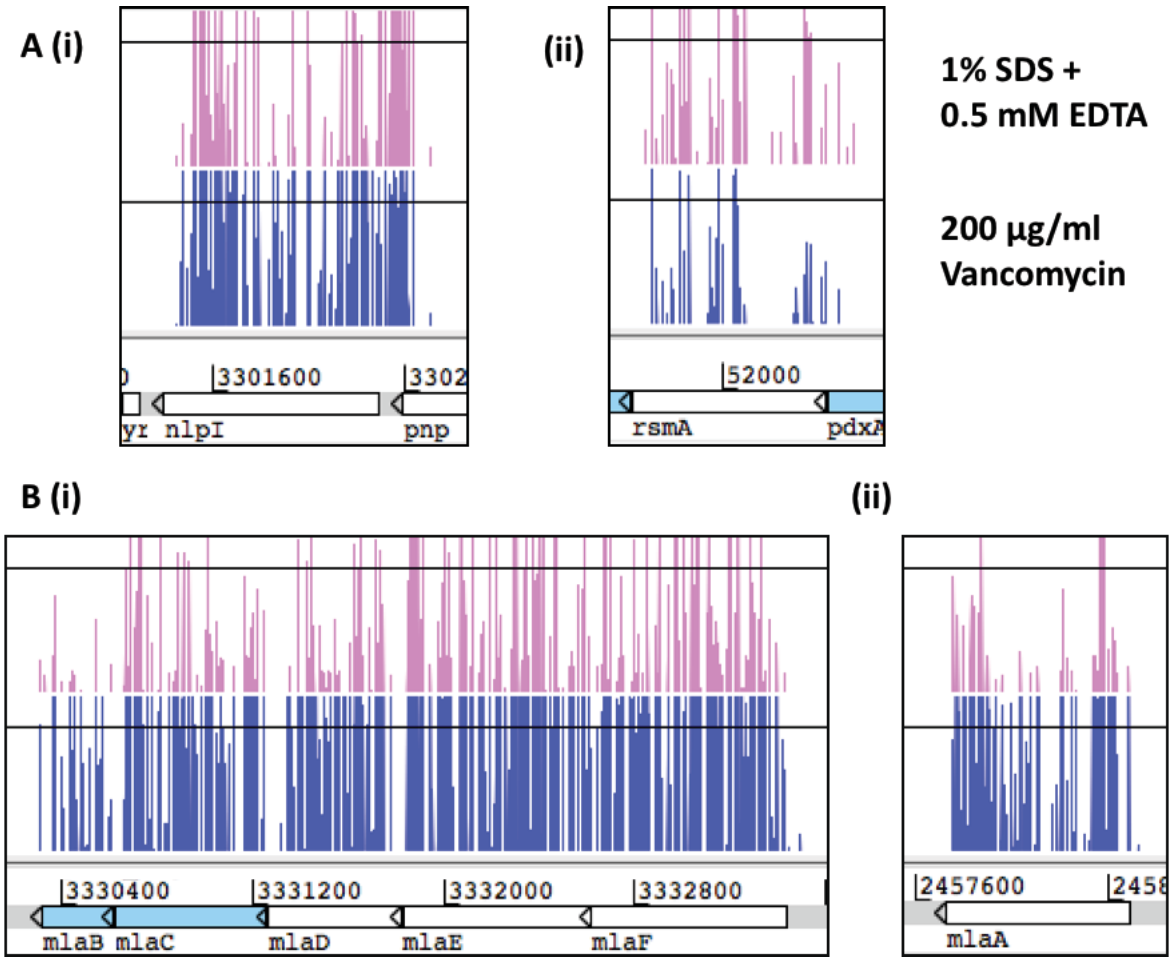

Figure 6.26 Transposon mutation sites that restore the sensitivity of a Δ*yhcB* mutant to vancomycin and SDS + EDTA

Transposon insertion data of the Δ*yhcB* library grown on LB agar plates supplemented with low concentrations of SDS + EDTA (pink) or vancomycin (blue). The frequency of mapped reads per insertion site were filtered to remove any unique insertions with fewer than 10 reads. The frequency of mapped reads (indicated by the insertion peaks above the gene track) was capped at 50. (A) The suppressor insertion data for the two genes (*nlpI* and *rsmA*) that, when disrupted, suppress a Δ*yhcB* mutation under all four conditions. Only the lowest concentration of each condition is presented here. (B) The suppressor insertion data for the *mla* operon (i) and *mlaA* (ii) for low concentrations of SDS + EDTA and vancomycin.

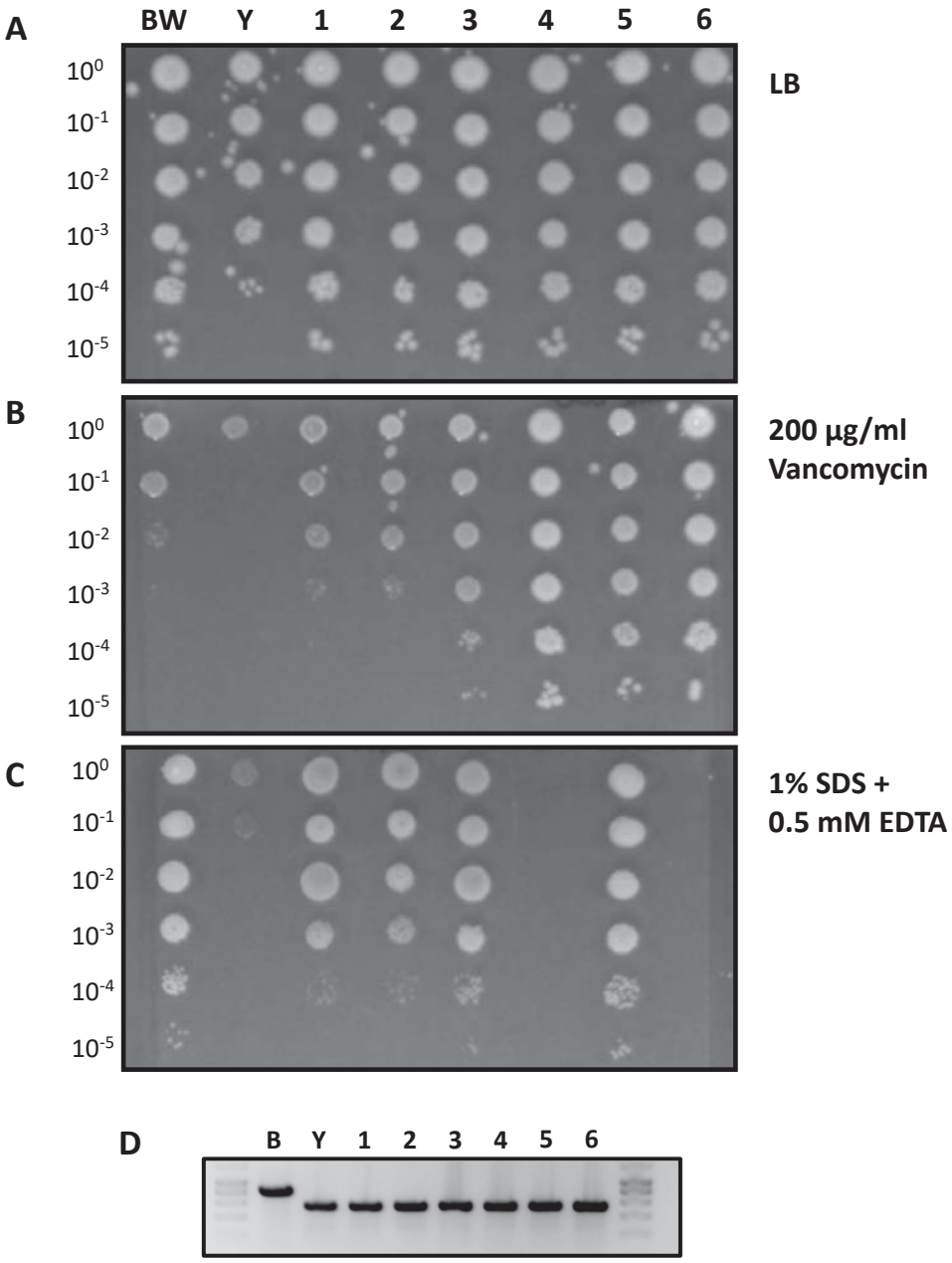

Figure 6.27 Validation of natural suppressor mutants

10-fold serial dilutions of the WT (BW), Δ*yhcB* (Y) and six natural suppressor mutants (1-6) isolated from a plate supplemented with 200 µg/ml vancomycin, grown on LB (A), LB supplemented with 200 µg/ml vancomycin (B) or 1% SDS + 0.5 mM EDTA (C). PCR amplification of the *yhcB* locus (D).

To investigate whether the suppressor mutations restore cell shape defects in addition to the barrier defects, the mutants were grown overnight in LB, diluted and visualised by microscopy. No cell measurements were taken, but all 6 mutants were visibly thinner than the Δ*yhcB* parent (Fig. 6.28 B-H). Mutants E to H (corresponding with mutants 3 to 6 above) were also visibly smaller, approaching the WT cell size (Fig. 6.28). To identify the mutation(s) that caused suppression of the *yhcB* deletion, the genomes of all 6 mutants were sequenced by microbesNG using Illumina technologies. The mutations identified by sequencing are presented in Table 6.1. Suppressor 1 has a non-synonymous single nucleotide polymorphism (SNP) within the gene *lpxC*, but the residue change is relatively close to the start of the gene and is hypothesised to alter the level of *lpxC* expression. The chromosome of suppressor 2 has undergone significant chromosome restructuring events that resulted in the separation of the *mla* operon from its promoter and is hypothesised to negate expression of the *mla* operon (equivalent to an operon-deletion mutant). Suppressor mutants number 3 and 5 both have single SNPs within *cdsA.* Both SNPs alter a different serine residue (in close proximity to each other in the protein sequence), and the overall change in both instances is the conversion of a neutral, polar residue to a non-polar, hydrophobic residue. Suppressor 3 also has an additional SNP in *puuD*. Finally, suppressors 4 and 6 have mutations at the same locus. Suppressor 4 has lost 2 residues (asparagine and phenylalanine) while suppressor 6 has gained 2 residues (asparagine and phenylalanine). The loss of asparagine and phenylalanine at this locus has already been characterised as the *mlaA** mutant. The *mlaA** mutant normally has a terminal phenotype and results in cell death caused by excessive blebbing (Sutterlin *et al*. 2016). The loss of these 2 residues from a membrane-imbedded helix is thought to alter the secondary structure allowing increased

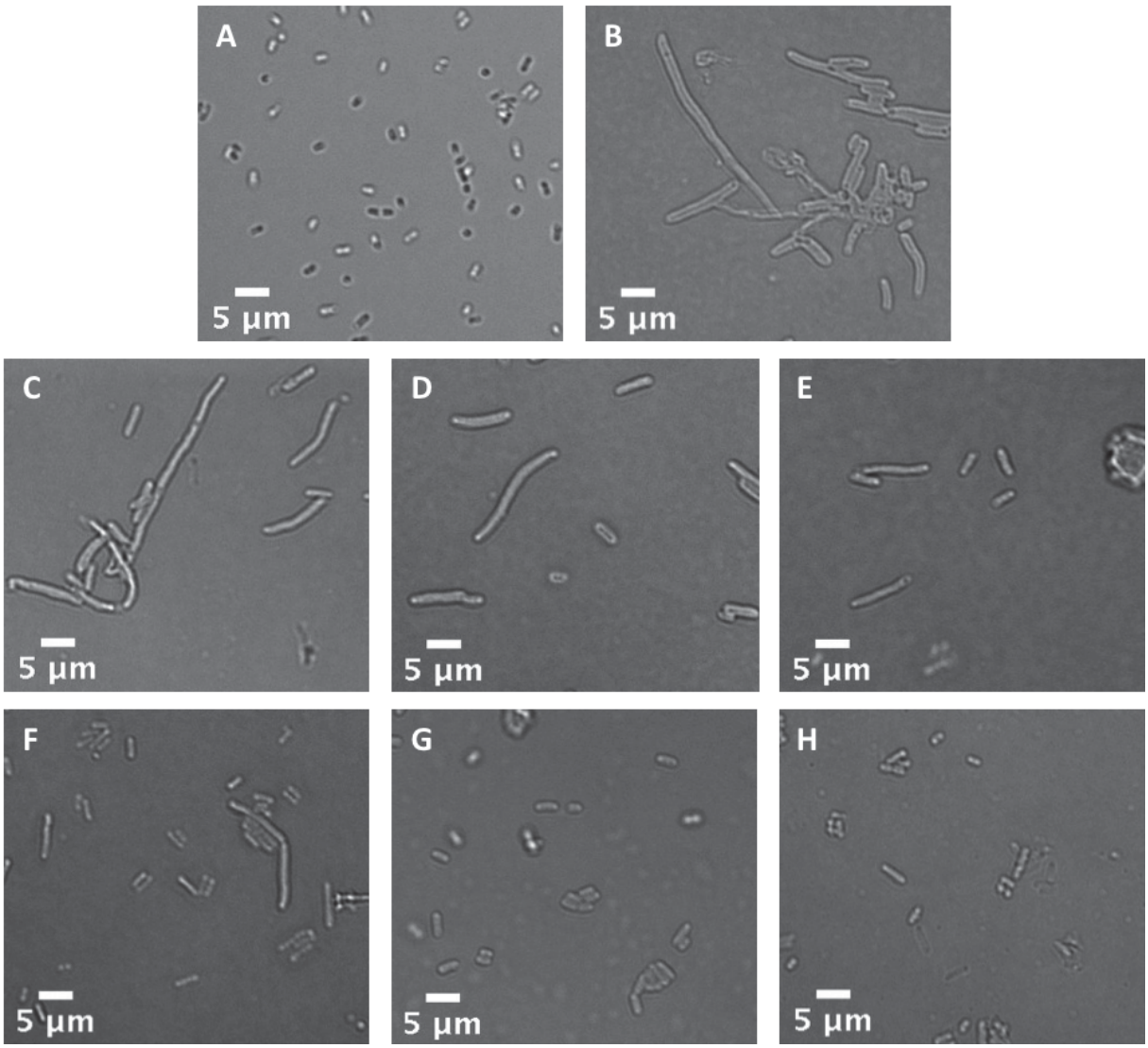

Figure 6.28 Microscopy of natural suppressor mutants isolated from plates supplemented with vancomycin

Visualisation by DIC microscopy of overnight cultures of cells grown in LB. (A) WT, (B) Δ*yhcB* and (C-H) six natural suppressor mutants isolated from a plate supplemented with 200 µg/ml vancomycin. A scale bar is shown for reference.

Table 6.1 Mutations within isolated *yhcB* natural suppressor mutants

Mutant	Change locus	Change type	Mutation	Effect
1	103,208	SNP	A>G	Non-synonymous mutation in *lpxC*
2	3,333,428	Inversion[a]		Separation of the mla operon from the operon promoters
2	3,333,440	Inversion[a]		
3	192,855	SNP	C>T	Serine>Leucine in *cdsA*
3	1,355,404	SNP	A>G	Isoleucine>Valine in *puuD*
4	2,458,352	Deletion	-TTGAAG	Conservative in frame deletion in *mlaA* → *mlaA**
5	192,878	SNP	A>G	Serine>Glycine in *cdsA*
6	2,458,352	Insertion	+TTGAAG	Conservative inframe insertion in *mlaA*

[a]The inversion locus is the output taken from BreSeq. Structural rearrangement is evident, but the precise reorganization cannot be determined from Illumina sequence data alone.

phospholipid access into the MlaA pore at this position. Given that the suppressor mutants 4 and 6 phenocopy, it is likely that the addition of 2 residues at this locus has the same effect as the *mlaA** mutation.

Further characterisation is required to fully understand the effects of each of these mutations, but LpxC is the first committed step in LPS biosynthesis, CdsA sits before the branch point in phospholipid biosynthesis and the *mla* pathway is involved in phospholipid trafficking. Together the data suggest a requirement for altered lipid biosynthesis and/or trafficking to restore the defect caused by deleting *yhcB*.

6.2.15 Comparison of lipid species between WT and ΔyhcB strains

To evaluate the phospholipid species of a *yhcB* mutant, total cell lipids were extracted using the Bligh & Dyer method (1959) from WT and Δ*yhcB* cells grown overnight. Phospholipids were then separated using thin layer chromatography with a chloroform:methanol:acetic acid solvent system at a ratio of 65:25:10. Three different volumes of each sample were loaded (Fig. 6.29). There was a visible additional spot in the sample from Δ*yhcB* cells ~1 cm above the origin. For the solvent system used, this spot is most likely to be lyso-PE or phosphatidic acid (Rowlett *et al.* 2017; Carpenter *et al.* 2014). In an *mlaD* mutant there is an increase in the amount of lyso-PE present and the phospholipid profile mirrors the one presented here for *yhcB* (Carpenter *et al.* 2014). Further analysis would be required to confirm the lipid species, but these data suggest that in a *yhcB* mutant there is either an increase in a breakdown product or an increase in precursor molecules.

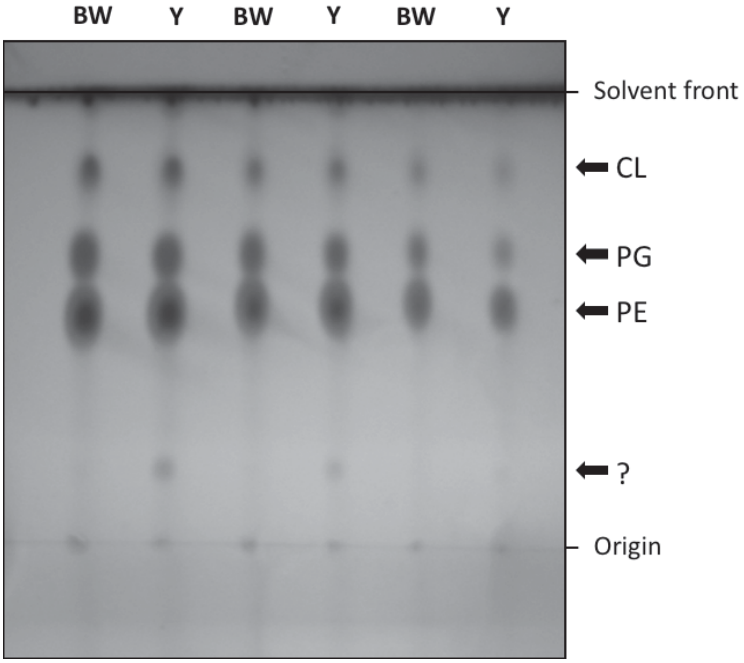

Figure 6.29 Thin layer chromatography separation of total cell phospholipids

Lipid species were extracted from WT (BW25113; BW) and Δ*yhcB* (Y) cells. Phospholipids were separated by thin layer chromatography using a chloroform:methanol:acetic acid solvent system in a ratio of 65:25:10. Based on the lipid separation presented by Carpenter *et al.* (2014) using the same solvent system, the major spots are assumed to be cardiolipin (CL), phosphatidyl glycerol (PG) and phosphatidyl ethanolamine (PE).

6.3 Discussion

This chapter presents extensive data to show that a *yhcB* mutant has a severe cell envelope defect. While the gene was indeed identified by eye from the initial polymyxin B TraDIS screen, the *yhcB* mutant was not identified as sensitive to polymyxin B based on statistical quantification of unique insertion sites. The findings of this chapter add further support to the conclusion that quantification of unique insertion sites can be insufficient to extract the subtle phenotypic effects observed in a sick mutant over ~5 generations of growth.

A role for YhcB remains to be defined. It is clear that a *yhcB* mutant has a defective cell envelope, but it is not clear which aspect of envelope biosynthesis is defective: whether the initial defect is in cell wall synthesis or in lipid synthesis. Wec mutants result in sequestration of available Und-P resulting in decreased availability of cell wall precursors. As a consequence, the available cell wall material is stretched further resulting in structurally less rigid cells with a larger cell size. The same could be true of a *yhcB* mutant; one hypothesis for the observed defects of *yhcB* is that cell wall synthesis is somehow impeded. In support of this, an *nlpI*-deletion mutant is the most prominent suppressor across all four OM stress conditions and deleting *nlpI* is reported to increase cell wall synthesis (Schwechheimer *et al.* 2015).

However, each of the identified natural suppressor mutations map to genes involved in phospholipid or LPS biosynthesis. While the mutations within LpxC and CdsA remain to be characterized there is already some available literature about the *mlaA** mutation. An *mlaA** mutant is thought to have increased flux of phospholipids from the OM inner leaflet to the OM outer leaflet and also has an increase in hepta-acylated LPS relative to WT (Sutterlin *et al.* 2016; May & Silhavy 2018). Both of these activities might serve to strengthen the OM in a Δ*yhcB* strain. Alternatively, the

hypervesiculation of an *mlaA** mutant induces shedding of OM material, which, although lethal in a WT strain, might suggest that a *yhcB* mutant has increased synthesis of OM lipids and can therefore tolerate shedding of the OM lipids via vesiculation. In addition, an *nlpI* mutant is also reported to hypervesiculate.

Furthermore, transposon insertion along the full length of the *fabF* CDS suppresses sensitivity to both vancomycin and SDS + EDTA. FabF is one of three β-ketoacyl-ACP synthases, but is the only one of the three that converts palmitoleate to *cis*-vaccenate. Deletion of *fabF* results in decreased amounts of cis-vaccenate and would shunt palmitoleate through other biosynthetic pathways (Garwin *et al.* 1980). Notably, palmitoyl-CoA is reported to inhibit FadR (fatty acid degradation regulon) binding to DNA. FadR is a transcriptional regulator that represses genes involved in fatty acid degradation, and activates genes involved in fatty acid synthesis. Palmitoyl-CoA is ~50-fold more efficient at inhibiting FadR activity than myristoyl-CoA (DiRusso *et al.* 1998). Therefore, it is possible that deletion of *fabF* might indirectly result in increased levels of Palmitoyl-CoA and increased inhibition of FadR. In support of this, insertions were identified along the full length of the *fadR* CDS in the SDS + EDTA suppressor data (but not vancomycin), suggesting deletion of *fadR* suppresses sensitivity to SDS + EDTA. These data indicate a requirement for reduced fatty acid synthesis or increased fatty acid degradation in a *yhcB* mutant.

Together, these data suggest that the balance of LPS and PGN synthesis in a *yhcB* mutant is uneven, and possibly weighted in favour of LPS biosynthesis, although this remains to be confirmed. Given that the LPS and PGN biosynthetic pathways share some initial biosynthetic precursors, it is possible that the flux of these precursors has been altered. In support of this, within the TraDIS insertion-suppressor data, a large number of insertions, for both the vancomycin and SDS +

EDTA datasets, map upstream of the promoter for *ispU* (Fig. 6.30). The gene *ispU* codes for undecaprenyl diphosphate synthase and the inserted transposons upstream of *ipsU* are restricted to one orientation, consistent with maintaining expression of *ispU*. Undecaprenyl diphosphate synthase catalyzes the synthesis of undecaprenyl diphosphate, which is the precursor to Und-P (Kato *et al*. 1999). These data suggest that altered levels of Und-P has a restorative effect in a *yhcB* deletion strain as the insertions suppress under both conditions of envelope stress. It cannot be determined from this data whether the transposon insertions upstream of *ispU* increase or decrease *ispU* expression. But overexpression of Und-P in a *wecE* mutant has a restorative effect as a result of increased cell wall synthesis, therefore it is more likely that these insertions increase *ispU* expression (Jorgenson *et al*. 2016).

YhcB has previously been reported to interact with proteins of the elongasome, and arguably a *yhcB* mutant has a cell-elongation defect when deleted. A schematic of these proteins is presented in figure 6.31 for reference and proteins with a reported positive bacterial two-hybrid interaction with YhcB are shown in colour. Surprisingly, the full length of RodZ was not conditionally essential in the *yhcB* TraDIS data. RodZ acts as a cross-membrane link that connects cell wall synthesis in the periplasm to the MreB filaments on the cytoplasmic face of the IM (Morgenstein *et al*. 2015). MreB is an actin homolog and categorized as a cytoskeletal protein. An *mreB* mutant adopts a round cell morphology, with a diameter roughly equivalent to the length of a cell. A RodZ mutant also has a round cell phenotype, but the diameter is roughly equivalent to the width of the parent rod-cell; RodZ mediates lateral cell wall growth. The *yhcB* TraDIS data suggest tethering of MreB to the IM (via the cytoplasmic domain of RodZ) is vital, but the coupling of MreB to cell wall synthesis via the periplasmic domain of

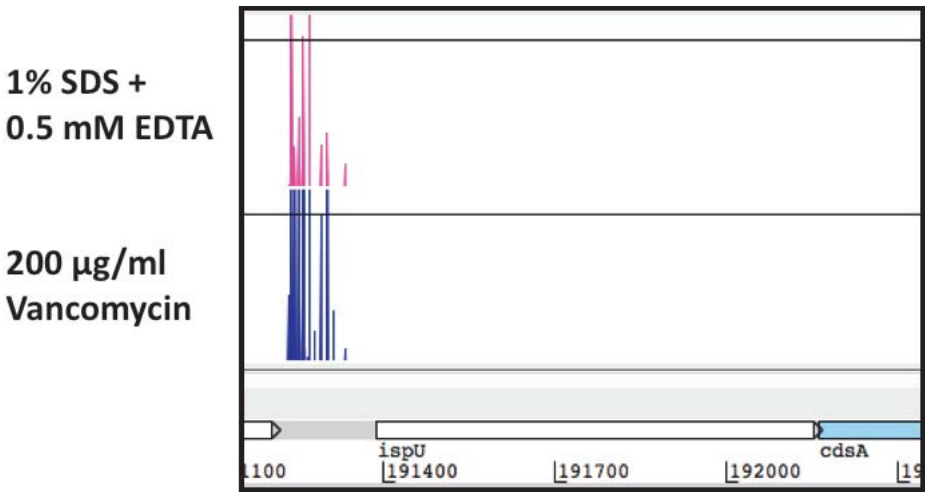

Figure 6.30 Transposon insertion sites upstream of *ispU* that suppress sensitivity to outer membrane stresses

Transposon insertion sites that confer resistance to 1% SDS + 0.5 mM EDTA (pink), and 200 µg/ml vancomycin (blue) in a Δ*yhcB* mutant map upstream of the gene *ispU*.

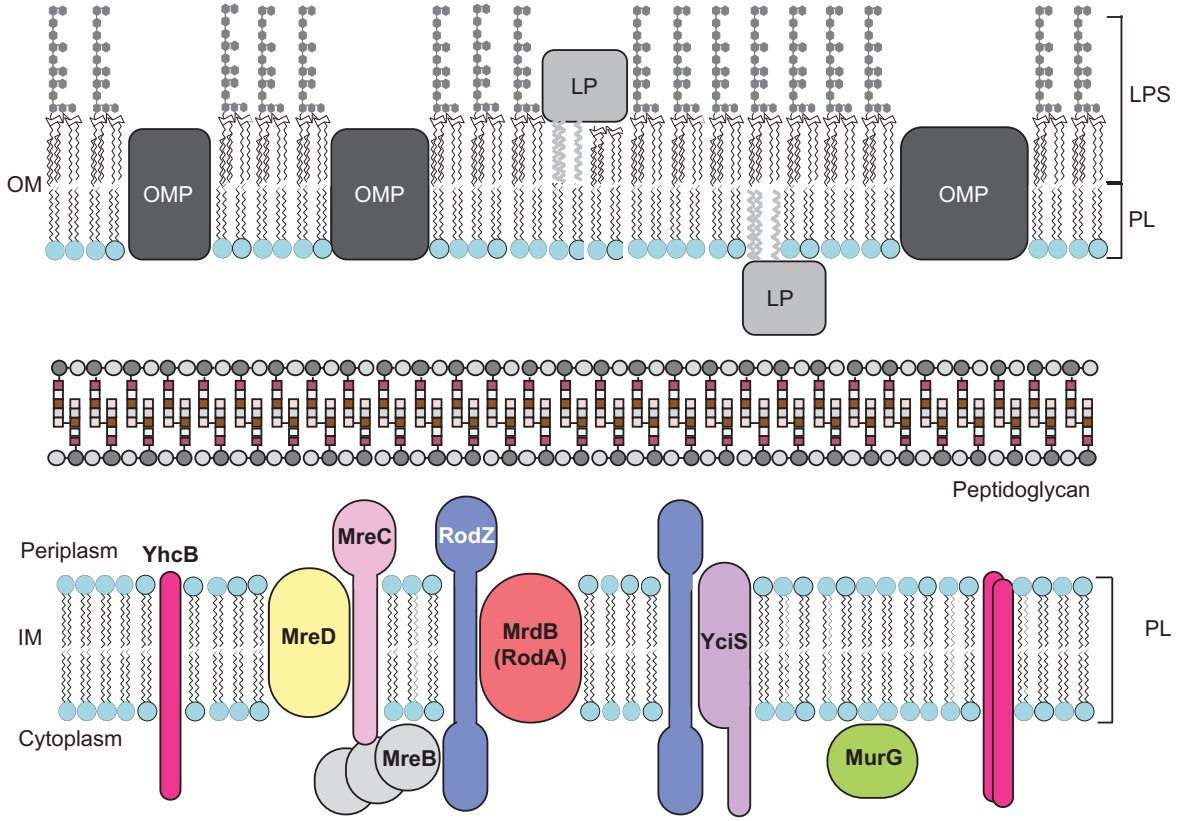

Figure 6.31 Cellular localisation of proteins reported to interact with YhcB

A schematic of the cell envelope. Proteins reported by Li *et al.* (2012) to have a positive interaction with YhcB in a bacterial two-hybrid screen are shown in colour. Known protein interactions within a complex are indicated by protein proximity.

RodZ is dispensable, possibly because a *rodZ* mutant has decreased lateral wall synthesis, which might be restorative in a *yhcB* mutant.

Further adding confusion to the picture is the selective conditional essentiality of enzymes with overlapping function, or in the same biosynthetic pathway. For example, why are MepM and MepS synthetically lethal but MepH is not? Why is BacA, but not BacB, BacC or BacD, synthetically lethal? And why is LdcA synthetically lethal, but no other enzymes of the PGN recycling pathway are? If cell wall synthesis or turnover was limited, you might expect AmpG, NagZ or AmpD to be synthetically lethal, but they are not (Uehara & Park 2008). The only thing that BacA and LdcA enzymes have in common is the release of D-alanine. However, D-alanine is also consumed in the L-alanine degradation pathway to synthesize pyruvate, but none of the genes *dadX*, *alr* or *dadA* are synthetically lethal with *yhcB*.

D-amino acids have been reported to induce physiological changes in *B. subtilis*, *P. aeruginosa* and *S. aureus*, although the precise mechanisms of action behind the morphological changes remain unknown (Kolodkin-Gal *et al*. 2010). In addition, peptidoglycan fragments have also been reported to act as cell signaling molecules (Shah *et al*. 2008). In *S. aureus*, a *pknB* (a serine/threonine kinase) mutant has significantly decreased amounts of the cell wall precursor UDP-MurNAc-Ala, which is hypothesized to be due to the positive regulation of MurC by *pknB* (Liebeke *et al*. 2010). Similarly, PknA from *C. glutamicum* has been reported to phosphorylate MurC *in vitro*, and the phosphorylation of YvcK in *B. subtilis* by PrkC is required for the correct localization of PBP1 (Fiuza *et al*. 2008; Foulquier *et al*. 2014). Altogether these findings infer a role for cell-wall metabolites in modulating cell wall synthesis via phosphorylation signal transduction mechanisms.

YhcB strongly self-interacts, and at least might form a dimer (indicated by the Western blot analysis presented in this chapter). Furthermore, the helical cytoplasmic domain was predicted to form a coiled-coil structure that closely resembles a HAMP domain. A HAMP domain is often found in transmembrane proteins involved in signaling and functions as a linker domain (Hulko *et al.* 2006; Jin & Inouye 1994). The HAMP acronym was derived from the proteins it is found in: histidine kinases, adenylyl cyclases, methyl-acceptors and phosphatases. The coding sequence of *yhcB* contains highly conserved histidine residues, as well as conserved serine and tyrosine residues, which are typically used by kinases in phospho-relay systems. It is possible that YhcB has a role in signal transduction pathways that regulate the availability of Und-P linked precursors involved in cell envelop synthesis.

Finally, the gene *yhcB* has been identified in a diverse range of large genomic screens. *yhcB* has been reported to belong to the *mazF* regulon in *E. coli*, *yhcB* is differentially regulated during space-flight in *S. enterica* Typhimurium, *yhcB* is upregulated in *S. enterica* Typhimurium at the mucosal wall of the chicken intestine, it is part of the colistin-resistome in *K. pneumoniae* and is required for the colonization of the light organ of the Hawaiian bobtail squid, *Euprymna scolopes*, by *Vibrio fischeri* (Brooks *et al.* 2014; Harvey *et al.* 2011; Jana *et al.* 2017; Wilson *et al.* 2007; Sauert *et al.* 2016). The diverse array of experimental evidence among a range of bacterial species that isolate a *yhcB* mutant highlights that, although *yhcB* is not essential, it appears to have a fundamental role in cell physiology. Furthermore, *yhcB* is highly conserved among gamma proteobacteria, and is present in 3 out of the 6 ESKAPE pathogens. Further characterization of *yhcB* would improve understanding of the bacterial cell envelope and might even present a novel antimicrobial target.

Chapter 7

Final Discussion

7.1 Summary

Overall, this thesis has examined the essential genome of *E. coli* K-12 and confirmed that discrepancies do arise when using different methodologies to classify gene essentiality, but close inspection of the data and comparison of the library both before and after growth in broth enabled characterisation of the major causes of discrepancies. The data revealed previously unreported essential genes and resolved previous contradictions of gene essentiality for some genes, e.g. *hda* and *holD*. The unprecedented resolution of the data also revealed boundaries between essential and non-essential domains at the bp level. This is a huge strength of TraDIS because not only did it enable identification of novel regulatory elements discussed in Chapter 4 but it has implications for the further use of TraDIS with appropriate transposon and experimental design.

Secondly, the transposon library characterised in Chapter 3 was used to identify genes that, when disrupted, confer sensitivity to polymyxin B. Analysis of this data confirmed existing reports of the sensitivity of certain mutants to membrane stress, validating the technique. Manual inspection of the data revealed two novel genes of unknown function that are potentially part of the polymyxin tolerome. However, the lack of identification of any conditionally-essential regulatory elements (as highlighted in Chapter 3 and 4) emphasises both the limitations of the insertion analysis used and the difficulties in elucidating a phenotype from manual inspection. Better analysis methods are required to garner the full wealth of information from a TraDIS dataset as discussed in Chapter 5.

Lastly, Chapter 6 centred around validating a gene of unknown function identified in the polymyxin TraDIS screen. However, a critical part of this analysis was the use of TraDIS to identify synthetically lethal relationships. This is an important strength of TraDIS data because the data can be used to inform genomic

interaction networks (Santa Maria *et al.* 2014), and can even reveal phenotypes for genes of unknown function with no previously available annotation data. It is plausible that, if applied to the Keio library of deletion mutants and coupled with literature reports, it might be possible to construct a whole-genome interaction network; a hypothesis previously presented in the GIANT coli study (Typas *et al.* 2008).

7.2 Additional uses of TraDIS

The uses of TraDIS presented in the introduction focused on the most common uses of TraDIS, however, TraDIS is a powerful tool and its use in answering biological questions is certainly not limited to just these three areas. For example, the identification of a single insertion site within GrpE that maps to a turn between 2 helices of the protein structure would suggest that TraDIS can actually be used to identify essential (or conditionally essential genes) that can function as split peptides, and furthermore reveals the borderline between functional domains. A finding that may be of use for downstream applications.

Other uses of TraDIS include combining transposon mutagenesis with cell sorting, termed "TraDISort" (Hassan *et al.* 2016; Paulsen *et al.* 2017). Flow cytometry coupled with TraDIS has been used to isolate mutants with a number of phenotypes. Kevin Young's groups combined the two to identify mutants with a larger cell volume (Jorgenson *et al.* 2016). Due to the larger cell size, these mutants were expected to have altered side-scatter and forward-scatter properties. The method was validated by their identification of the Wec mutants. However, their study was limited as the mutant pool they tested comprised only ~5,000 mutants. The method can also be adapted to sort cells based on fluorescent properties. The first study to use

fluorescence-activated cell sorting (FACS) separated differentially fluorescent *A. baumannii* cells stained with ethidium bromide (Hassan *et al.* 2016; Paulsen *et al.* 2017). Ethidium bromide is a common substrate of efflux pumps and has a different fluorescence profile inside and outside of the cell. Mutants with defective efflux systems would therefore fluoresce differently to those able to export ethidium bromide. These mutants were quantified by flow cytometry and enriched for via cell sorting. The populations of differential ethidium bromide fluorescence were then sequenced to identify the disrupted genes that confer either reduced efflux or increased permeability.

A conceivable use of TraDISort could be to identify protein interaction partners. For example, it might be possible to construct a chromosomal mutant with a fluorescent protein (FP) fragment tagged onto a gene of interest, which, when translated will code for a functional protein with a fused FP peptide. A construct harbouring a transposon that carries the remainder of the FP could then be introduced into the bacterium. This would result in the synthesis of a fusion peptide when inserted into a CDS. In theory, if the transposon has inserted into a gene whose product interacts with the original protein of interest, this would bring the two fragments of the FP together to produce a functional FP, which would be detectable using FACS. This gives you a genome scale view of protein-protein interaction, which has not been possible before. A similar use of a transposon to identify protein localisation is the use of Tn*phoA*. The gene *phoA* encodes an alkaline phosphatase that is only functional in the periplasm. Tn*phoA* can therefore inform which cellular compartment a protein is in and also the topology of integral membrane proteins (Manoil & Beckwith 1985).

Lastly, it is also possible to use this technique to probe chromosomal regulatory elements (Christen *et al.* 2014). For example, transposon libraries that make use of an inducible outward-facing promoter, although primarily used in drug-resistance screens (Meredith *et al.* 2012), can also be used to understand native transcription expression levels and the to identify the position of essential promoters. Furthermore, appropriate design of a transposon might enable identification of every actively translated CDS of the genome. This is an important and functionally relevant screen given the reports that small ORFs are classically overlooked, suggesting a number of small genes remain to be identified, and a limitation in genome annotation is the validation of bioinformatically assigned annotations. Furthermore, existing analyses to identify whole-genome coding sequences, such as RBS footprinting, are not a direct measure of translational activity as these data are a measure of RBS binding and not ribosomal activity.

7.3 Limitations

TraDIS is a powerful tool but is limited by the ability to construct a transposon library in your strain of choice. With the development of molecular genetics tools, it is becoming increasingly tangible to construct large mutant libraries for some of these strains. For example, it is reportedly difficult to construct chromosomal mutants within the *Fusobacterium* species, but a method for genetic manipulation has recently been published (Casasanta *et al.* 2017). It is conceivable that in time, perhaps improved methods might facilitate the construction of transposon libraries within *Fusobacterium*. Improved methodologies are also useful for species with existing transposon libraries. For example, an enhanced method was published for the construction of a dense transposon library in *S. aureus* that no longer necessitates

growth at a higher temperature to remove the transposon-delivery plasmid, which limited the characterisation of temperature sensitive strains (Santiago *et al.* 2015). However, despite technique advances, even if it is possible to construct a transposon library in your species of interest, working with a particular strain can be precluded by intrinsic antibiotic resistance. Strains with high levels of intrinsic resistance limit the choices of antibiotic resistance cassette that can be used in the transposon, which is especially challenging if you wish to construct a transposon library in a multi drug resistant isolate. Finally, even if construction of a transposon library is possible, there are a number of underlying mechanisms that can lower the frequency of transposon insertion resulting in the false classification of essential genes. For example, occlusion by DNA-binding proteins, regions of extreme chromosomal structure and proximity to the chromosome terminus, among others discussed in Chapter 3.

7.4 Concluding remarks

However, despite these challenges, TraDIS remains a powerful tool with a wide range of applications for a diverse range of species, which will only be expanded in the coming years. Furthermore, the wealth of information available from a single, standard TraDIS essentiality screen will likely improve too. Even a highly saturated transposon library within the model organism *E. coli* K-12 contained transposon-free regions that remain to be fully characterised. The prevalence of intragenic IFRs in this transposon library were especially surprising, and the frequency distribution of IFR sizes was indicative of further underlying biological phenomenon that remain to be characterised. TraDIS may yet reveal new and unexpected features of chromosome biology.

Appendix 3.1. Comparison of essential genes identified by Keio, PEC and TraDIS

TraDIS only	Keio only	PEC only (W3110)	TraDIS-Keio	TraDIS-PEC	Keio-PEC	All 3
aceF	*alsK*	*argU*	*cydA*	*alaS*	*degS*	*lptD (imp)*
cydB	*bcsB*	*argX*	*cydC*	*coaA*	*folK*	*erpA (yadR)*
cydD	*mazE (chpR)*	*cysT*	*dicA*	*coaE*	*ftsE*	*bamA (yaeT)*
cydX (ybgT)[a]	*chpS*	*efp*	*purB*	*dnaG*	*ftsK*	*lptE (rlpB)*
dapF	*entD*	*ffs*	*racR*	*dnaT*	*ftsN*	*murJ (mviN)*
dcd	*minD*	*glyT*	*rpoE*	*folB*	*ftsX*	*prs (prsA)*
fabH	*minE*	*hisR*	*tadA*	*glmM*	*ribB*	*tsaB (yeaZ)*
fdx	*waaU (rfaK)*	*kdsC*	*ubiB*	*glyS*	*rne*	*bamD (yfiO)*
folP	*rnc*	*leuU*	*ubiD*	*groL*	*secD*	*nadK (yfjB)*
glyA	*tdcF*	*leuW*	*wzyE*	*hda*	*secF*	*tsaD (ygjD)*
guaA	*tnaB*	*leuZ*	*cohE (ymfK)*	*ileS*	*secM*	*lptA (yhbN)*
hemE	***yabQ***[bcd]	*polA*		*nusB*	*spoT*	*tsaC (yrdC)*
higA	*yafF*[c]	*priA*		*parC*	*rseP (yaeL)*	*waaA (kdtA)*
hipB	*yagG*	*proM*		*prfB*	*yceQ*	*tsaE (yjeE)*
holD	*ydfB*	*serT*		*rho*	*yejM*	*lptF (yjgP)*
hscA	*ydiL*	*serV*		*rpoD*	*lptC (yrbK)*	*lptG (yjgQ)*
ihfA	*yefM*	*thrU*		*rsgA*		*ribF*
iraM	*mqsA (ygiT)*	*trpT*		*lptB (yhbG)*		*lspA*
iscS	*yhbV*					*ispH*
iscU	*yhhQ*					*dapB*
lipA	***yibJ***[bce]					*folA*
lpd	*ubiJ (yigP)*					*ftsL*
lpxL	***yqgD***[bf]					*ftsI*
lysS	*rsmI (yraL)*					*murE*
mnmA	*mlaB (yrbB)*					*murF*
pdxH						*mraY*
pheM						*murD*
priB						*ftsW*
ptsI						*murG*
rbfA						*murC*
relB						*ftsQ*
rimM						*ftsA*
rluD						*ftsZ*
rnt						*lpxC*
rpe						*secA*
rplA						*can*
rplK						*hemL*
rplY						*dapD*
rpmF						*map*
rpmI						*rpsB*
rpsF						*tsf*
rpsO						*pyrH*
rpsT						*frr*
rpsU						*dxr*
safA						*ispU*

sucA
sucB
thyA
tktA
tonB
trpL
ttcC
tusE (yccK)
ubiE
ubiG
ubiH
ubiX
ybeY
ycaR
yciS
ydaE
ydaS
ydcD
yddL
ydfO
ydhL
yedN
yffS
ygeF
ygeG
ygeN
ygfZ
yjbS
ykfM
ymfE
ymiB
ynbG
yncH
yobI
yqcG
yqeL

cdsA
lpxD
fabZ
lpxA
lpxB
dnaE
accA
tilS
proS
hemB
ribD
ribE
thiL
dxs
ispA
dnaX
adk
hemH
lpxH
cysS
folD
mrdB
mrdA
nadD
holA
leuS
lnt
glnS
fldA
infA
lolA
serS
rpsA
msbA
lpxK
kdsB
mukF
mukE
mukB
asnS
fabA
fabD
fabG
acpP
tmk
holB
lolC
lolD
lolE
pth
ispE
lolB

hemA
prfA
prmC
kdsA
topA
ribA
fabI
tyrS
ribC
pheT
pheS
rplT
infC
thrS
nadE
gapA
aspS
argS
pgsA
metG
folE
gyrA
nrdA
nrdB
folC
accD
fabB
gltX
ligA
zipA
dapE
dapA
der
hisS
ispG
suhB
acpS
era
lepB
pssA
rplS
trmD
rpsP
ffh
grpE
csrA
ispF
ispD
ftsB
eno
pyrG
lgt

fbaA
pgk
metK
yqgF
plsC
parE
cca
infB
nusA
ftsH
obgE
rpmA
rplU
ispB
murA
rpsI
rplM
mreD
mreC
mreB
accB
accC
def
fmt
rplQ
rpoA
rpsD
rpsK
rpsM
secY
rplO
rpmD
rpsE
rplR
rplF
rpsH
rpsN
rplE
rplX
rplN
rpsQ
rpmC
rplP
rpsC
rplV
rpsS
rplB
rplW
rplD
rplC
rpsJ
fusA

	rpsG
	rpsL
	trpS
	yrfF
	asd
	rpoH
	ftsY
	glyQ
	gpsA
	coaD
	rpmB
	dfp
	dut
	gmk
	gyrB
	dnaN
	dnaA
	rpmH
	rnpA
	yidC
	glmS
	glmU
	hemD
	hemC
	hemG
	yihA
	murI
	murB
	birA
	secE
	nusG
	rplJ
	rplL
	rpoB
	rpoC
	ubiA
	plsB
	lexA
	dnaB
	ssb
	groS
	psd
	orn
	rpsR
	ppa
	valS
	dnaC

[a]BW25113 (CP009273.1) gene names have been used unless otherwise specified. Alternative gene names used by Keio or PEC are shown in brackets

[b]The Keio naming convention is used for all genes in bold

[c]The annotation of this gene varies between the Keio collection and BW25113 (CP009273.1)

[d]*yabQ* is not annotated in BW25113 (CP009273.1), but corresponds with the 2nd half of *yabP*

[e]*yibJ* is annotated as *rhsJ* in BW25113 (CP009273.1), but the annotation of *rhsJ* extends beyond the first stop codon

[f]*yqgD* is not annotated in BW25113 (CP009273.1)

Appendix 3.2. Causes of discrepancies between datasets

Gene	Venn group	Cause of discrepancy
alsK	K	Errors in library construction
bcsB	K	Errors in library construction
chpS	K	Anti-Toxin
entD	K	Errors in library construction
mazE (chpR)	K	Anti-Toxin
minD	K	Errors in library construction
minE	K	Errors in library construction
mlaB (yrbB)	K	Errors in library construction
mqsA (ygiT)	K	Genes containing a transposon free region
rnc	K	Polar insertions
rsmI (yraL)	K	Errors in library construction
tdcF	K	Unclear
tnaB	K	Errors in library construction
ubiJ (yigP)	K	Polar insertions
waaU (rfaK)	K	Genes containing a transposon free region
yabQ	K	Genes containing a transposon free region
yafF	K	Genes containing a transposon free region
yagG	K	Unclear
ydfB	K	Unclear
ydiL	K	Unclear
yefM	K	Anti-Toxin
yhbV	K	Unclear
yhhQ	K	Errors in library construction
yibJ	K	Genes containing a transposon free region
yqgD	K	Genes containing a transposon free region
degS	KP	Conditionally essential
folK	KP	Conditionally essential
ftsE	KP	Errors in library construction
ftsK	KP	Genes containing a transposon free region
ftsN	KP	Genes containing a transposon free region
ftsX	KP	Genes containing a transposon free region
lptC (yrbK)	KP	Genes containing a transposon free region
ribB	KP	Genes containing a transposon free region
rne	KP	Genes containing a transposon free region
rseP (yaeL)	KP	Conditionally essential
secD	KP	Errors in library construction
secF	KP	Errors in library construction
secM	KP	Genes containing a transposon free region
spoT	KP	Genes containing a transposon free region
yceQ	KP	Polar insertions
yejM	KP	Genes containing a transposon free region
cohE (ymfK)	KT	Unclear
cydA	KT	Errors in library construction

cydC	KT	Errors in library construction
dicA	KT	Errors in library construction
purB	KT	Conditionally essential
racR	KT	Unclear
rpoE	KT	Errors in library construction
tadA	KT	Errors in library construction
ubiB	KT	Conditionally essential
ubiD	KT	Conditionally essential
wzyE	KT	Unclear
argU	P	RNA genes not considered in our analysis
argX	P	RNA genes not considered in our analysis
cysT	P	RNA genes not considered in our analysis
efp	P	Errors in library construction
ffs	P	RNA genes not considered in our analysis
glyT	P	RNA genes not considered in our analysis
hisR	P	RNA genes not considered in our analysis
kdsC	P	Polar insertions
leuU	P	RNA genes not considered in our analysis
leuW	P	RNA genes not considered in our analysis
leuZ	P	RNA genes not considered in our analysis
polA	P	Genes containing a transposon free region
priA	P	Errors in library construction
proM	P	RNA genes not considered in our analysis
serT	P	RNA genes not considered in our analysis
serV	P	RNA genes not considered in our analysis
thrU	P	RNA genes not considered in our analysis
trpT	P	RNA genes not considered in our analysis
alaS	PT	Errors in library construction
coaA	PT	Errors in library construction
coaE	PT	Errors in library construction
dnaG	PT	Errors in library construction
dnaT	PT	Unclear
folB	PT	Conditionally essential
glmM	PT	Errors in library construction
glyS	PT	Errors in library construction
groL	PT	Errors in library construction
hda	PT	Errors in library construction
ileS	PT	Errors in library construction
lptB (yhbG)	PT	Errors in library construction
nusB	PT	Polar insertions
parC	PT	Errors in library construction
prfB	PT	Errors in library construction
rho	PT	Errors in library construction
rpoD	PT	Errors in library construction
rsgA	PT	Errors in library construction

Appendix 5.1. COG groups of the conditionally essential genes required for growth in sub-MIC concentrations of polymyxin B

Gene	COG	Function
ackA	C	Catalyzes the formation of acetyl phosphate from acetate and ATP. Can also catalyze the reverse reaction (By similarity)
fdx	C	Ferredoxin, 2FE-2S
ndh	C	NADH dehydrogenase
pta	C	Involved in acetate metabolism (By similarity)
ftsB	D	Essential cell division protein. May link together the upstream cell division proteins, which are predominantly cytoplasmic, with the downstream cell division proteins, which are predominantly periplasmic (By similarity)
ftsL	D	Essential cell division protein. May link together the upstream cell division proteins, which are predominantly cytoplasmic, with the downstream cell division proteins, which are predominantly periplasmic (By similarity)
dapF	E	Catalyzes the stereoinversion of LL-2,6- diaminoheptanedioate (L,L-DAP) to meso-diaminoheptanedioate (meso- DAP), a precursor of L-lysine and an essential component of the bacterial peptidoglycan (By similarity)
gmhB	E	d,d-heptose 1,7-bisphosphate phosphatase
mtn	F	Catalyzes the irreversible cleavage of the glycosidic bond in both 5'-methylthioadenosine (MTA) and S- adenosylhomocysteine (SAH AdoHcy) to adenine and the corresponding thioribose, 5'-methylthioribose and S-ribosylhomocysteine, respectively (By similarity)
pgm	G	Phosphoglucomutase (EC 5.4.2.2
rpiA	G	phosphoriboisomerase A
tktA	G	Transketolase (EC 2.2.1.1)
ratA	I	Cyclase dehydrase
rbfA	J	Associates with free 30S ribosomal subunits (but not with 30S subunits that are part of 70S ribosomes or polysomes). Essential for efficient processing of 16S rRNA. May interact with the 5'-terminal helix region of 16S rRNA (By similarity)
rne	J	Ribonuclease E
rnpA	J	RNaseP catalyzes the removal of the 5'-leader sequence from pre-tRNA to produce the mature 5'-terminus. It can also cleave other RNA substrates such as 4.5S RNA. The protein component plays an auxiliary but essential role in vivo by binding to the 5'-leader sequence and broadening the substrate specificity of the ribozyme (By similarity)
rpmE	J	Binds the 23S rRNA (By similarity)
rpmF	J	50s ribosomal protein l32
ychF	J	gtp-binding protein
rnc	K	Digests double-stranded RNA. Involved in the processing of primary rRNA transcript to yield the immediate precursors to the large and small rRNAs (23S and 16S). Also processes some mRNAs, and tRNAs when they are encoded in the rRNA operon (By similarity)
rpoS	K	Sigma factors are initiation factors that promote the attachment of RNA polymerase to specific initiation sites and are then released (By similarity)
ydfK	K	DNA-binding, transcriptional regulator
ynaE	K	DNA-binding, transcriptional regulator
dnaQ	L	dna polymerase iii epsilon subunit
holC	L	DNA polymerase iii, chi subunit
nudB	L	dATP pyrophosphohydrolase
priA	L	Primosomal protein n'
rnhA	L	Endonuclease that specifically degrades the RNA of RNA- DNA hybrids (By similarity)
bamB	M	Part of the outer membrane protein assembly complex, which is involved in assembly and insertion of beta-barrel proteins into the outer membrane (By similarity)

galE	M	udp-glucose 4-epimerase
galU	M	UTP-glucose-1-phosphate uridylyltransferase
hldD	M	Catalyzes the interconversion between ADP-D-glycero- beta-D-manno-heptose and ADP-L-glycero-beta-D-manno-heptose via an epimerization at carbon 6 of the heptose (By similarity)
hldE	M	Catalyzes the ADP transfer to D-glycero-D-manno-heptose 1-phosphate, yielding ADP-D,D-heptose (By similarity)
lpp	M	major outer membrane lipoprotein
mreC	M	Involved in formation and maintenance of cell shape (By similarity)
waaB	M	Glycosyl transferase (Group 1
waaC	M	lipopolysaccharide heptosyltransferase i
waaF	M	heptosyltransferase ii
waaP	M	(lipo)polysaccharide
waaR	M	Glycosyl Transferase
dnaK	O	Acts as a chaperone (By similarity)
glnD	O	Modifies, by uridylylation or deuridylylation the PII (GlnB) regulatory protein (By similarity)
grpE	O	Participates actively in the response to hyperosmotic and heat shock by preventing the aggregation of stress-denatured proteins, in association with DnaK and GrpE. It is the nucleotide exchange factor for DnaK and may function as a thermosensor. Unfolded proteins bind initially to DnaJ
smpB	O	Binds specifically to the SsrA RNA (tmRNA) and is required for stable association of SsrA with ribosomes (By similarity)
surA	O	Chaperone involved in the correct folding and assembly of outer membrane proteins. Recognizes specific patterns of aromatic residues and the orientation of their side chains, which are found more frequently in integral outer membrane proteins. May act in both early periplasmic and late outer membrane-associated steps of protein maturation (By similarity)
nhaA	P	Na(2+) H(+) antiporter that extrudes sodium in exchange for external protons (By similarity)
tusD	P	Part of a sulfur-relay system required for 2-thiolation of 5-methylaminomethyl-2-thiouridine (mnm(5)s(2)U) at tRNA wobble positions. Accepts sulfur from TusA and transfers it in turn to TusE (By similarity)
rimP	S	Required for maturation of 30S ribosomal subunits (By similarity)
ydfA	S	Protein of unknown function (DUF1391)
ygfZ	S	Folate-binding protein involved in regulating the level of ATP-DnaA and in the modification of some tRNAs. It is probably a key factor in regulatory networks that act via tRNA modification, such as initiation of chromosomal replication (By similarity)
yqeJ	S	NA
arcA	T	response regulator
dksA	T	Transcription factor that acts by binding directly to the RNA polymerase (RNAP). Required for negative regulation of rRNA expression and positive regulation of several amino acid biosynthesis promoters
iraM	T	Inhibits RpoS proteolysis by regulating RssB activity, thereby increasing the stability of the sigma stress factor RpoS during magnesium starvation (By similarity)
rcsB	T	regulatoR
rcsC	T	hybrid sensory kinase in two-component regulatory system with RcsB and YojN
rcsD	T	Phosphotransfer intermediate protein in two-component regulatory system with RcsBC
rseA	T	sigma-e factor negative regulatory protein
secF	U	Part of the Sec protein translocase complex. Interacts with the SecYEG preprotein conducting channel. SecDF uses the proton motive force (PMF) to complete protein translocation after the ATP-dependent function of SecA (By similarity)
cydC	V	(ABC) transporter
sapD	V	Peptide ABC transporter ATP-binding protein

tpr	-
trpL	-
yddK	-
ymiB	-
yobI	-
yqcG	-
yqeL	-

Appendix 6.1. Enriched GO groups within the *yhcB* conditionally essential genes

GO biological process complete	*E. coli* (4306)[a]	*yhcB* (104)[b]	EP[c]	FE[d]	raw P-value	FDR
protein import (GO:0017038)	4	3	0.1	31.05	4.16E-04	3.74E-02
establishment of protein localization (GO:0045184)	39	6	0.94	6.37	5.75E-04	4.73E-02
bacteriocin transport (GO:0043213)	6	4	0.14	27.6	5.50E-05	8.58E-03
enterobacterial common antigen biosynthetic process (GO:0009246)	10	4	0.24	16.56	2.44E-04	2.41E-02
enterobacterial common antigen metabolic process (GO:0046378)	10	4	0.24	16.56	2.44E-04	2.33E-02
cellular polysaccharide biosynthetic process (GO:0033692)	69	13	1.67	7.8	3.78E-08	1.12E-04
polysaccharide biosynthetic process (GO:0000271)	70	13	1.69	7.69	4.40E-08	6.52E-05
cellular carbohydrate biosynthetic process (GO:0034637)	72	13	1.74	7.48	5.91E-08	5.83E-05
cellular polysaccharide metabolic process (GO:0044264)	76	13	1.84	7.08	1.04E-07	7.70E-05
polysaccharide metabolic process (GO:0005976)	79	13	1.91	6.81	1.56E-07	7.69E-05
carbohydrate biosynthetic process (GO:0016051)	93	14	2.25	6.23	1.35E-07	7.98E-05
cellular carbohydrate metabolic process (GO:0044262)	131	15	3.16	4.74	1.16E-06	4.91E-04
carbohydrate derivative biosynthetic process (GO:1901137)	178	16	4.3	3.72	9.21E-06	2.48E-03
carbohydrate derivative metabolic process (GO:1901135)	288	21	6.96	3.02	7.02E-06	2.08E-03
cellular macromolecule biosynthetic process (GO:0034645)	348	24	8.41	2.86	3.39E-06	1.26E-03
macromolecule biosynthetic process (GO:0009059)	351	24	8.48	2.83	3.91E-06	1.29E-03
carbohydrate metabolic process (GO:0005975)	245	16	5.92	2.7	3.38E-04	3.13E-02
macromolecule metabolic process (GO:0043170)	779	38	18.81	2.02	1.00E-05	2.48E-03
cellular biosynthetic process (GO:0044249)	746	36	18.02	2	3.77E-05	7.99E-03
organic substance biosynthetic process (GO:1901576)	756	36	18.26	1.97	4.19E-05	8.28E-03
cellular macromolecule metabolic process (GO:0044260)	652	31	15.75	1.97	1.68E-04	1.84E-02
biosynthetic process (GO:0009058)	762	36	18.4	1.96	4.58E-05	8.48E-03
primary metabolic process (GO:0044238)	1258	49	30.38	1.61	1.84E-04	1.88E-02
organic substance metabolic process (GO:0071704)	1526	56	36.86	1.52	1.71E-04	1.81E-02
cellular process (GO:0009987)	2047	69	49.44	1.4	1.47E-04	1.67E-02
lipopolysaccharide core region metabolic process (GO:0046401)	25	6	0.6	9.94	6.79E-05	9.58E-03
lipopolysaccharide core region biosynthetic process (GO:0009244)	25	6	0.6	9.94	6.79E-05	9.15E-03
lipopolysaccharide biosynthetic process (GO:0009103)	37	7	0.89	7.83	6.19E-05	9.18E-03
lipopolysaccharide metabolic process (GO:0008653)	39	7	0.94	7.43	8.32E-05	1.07E-02
oligosaccharide biosynthetic process (GO:0009312)	39	7	0.94	7.43	8.32E-05	1.03E-02
liposaccharide metabolic process (GO:1903509)	50	8	1.21	6.62	5.27E-05	9.19E-03
oligosaccharide metabolic process (GO:0009311)	50	8	1.21	6.62	5.27E-05	8.68E-03
lipid biosynthetic process (GO:0008610)	95	11	2.29	4.79	3.17E-05	7.22E-03
cellular lipid metabolic process (GO:0044255)	132	11	3.19	3.45	4.75E-04	4.14E-02
lipid metabolic process (GO:0006629)	135	11	3.26	3.37	5.68E-04	4.81E-02
cell division (GO:0051301)	54	8	1.3	6.13	8.60E-05	1.02E-02

[a]GO groups of the whole genome of *E. coli*
[b]GO groups of the conditionally essential genes in the Δ*yhcB* transposon library
[c]Expected proportion

REFERENCES

Abdelraouf, K. *et al.*, (2012). Characterization of polymyxin B-induced nephrotoxicity: implications for dosing regimen design. Antimicrobial Agents and Chemotherapy, 56(9), pp.4625-4629.

Abellón-Ruiz, J. *et al.*, (2017). Structural basis for maintenance of bacterial outer membrane lipid asymmetry. Nature Microbiology, 2(12), pp.1616-1623.

Abraham, E.P. & Chain, E., (1940). An enzyme from bacteria able to destroy penicillin. Reviews of Infectious Diseases, 10(4), pp.677-8.

Ades, S.E. *et al.*, (1999). The *Escherichia coli* sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. Genes & Development, 13(18), pp.2449-61.

Ainsworth, G.C., Brown, A.M. & Brownlee, G., (1947). 'Aerosporin', an antibiotic produced by *Bacillus aerosporus* Greer. Nature, 160(4060), pp.263-263.

Akerley, B.J. *et al.*, (2002). A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. Proceedings of the National Academy of Sciences of the United States of America, 99(2), pp.966-71.

Akerley, B.J. *et al.*, (1998). Systematic identification of essential genes by in vitro mariner mutagenesis. Proceedings of the National Academy of Sciences of the United States of America, 95(15), pp.8927-32.

Alba, B.M. *et al.*, (2001). *degS* (*hhoB*) is an essential *Escherichia coli* gene whose indispensable function is to provide sigma (E) activity. Molecular Microbiology, 40(6), pp.1323-33.

Anderson, M.S. & Raetz, C.R., (1987). Biosynthesis of lipid A precursors in *Escherichia coli*. A cytoplasmic acyltransferase that converts UDP-*N*-acetylglucosamine to UDP-3-O-(R-3-hydroxymyristoyl)-*N*-acetylglucosamine. The Journal of Biological Chemistry, 262(11), pp.5159-69.

Andrews, J.M., (2001). Determination of minimum inhibitory concentrations. The Journal of Antimicrobial Chemotherapy, 48 Suppl. 1, pp.5-16.

Anon, (1945). Penicillin's finder assays its future; Sir Alexander Fleming says improved dosage method is needed to extend use. The New York Times, p.21.

Ashburner, M. *et al.*, (2000). Gene Ontology: tool for the unification of biology. Nature Genetics, 25(1), pp.25-29.

Ashkenazy, H. *et al.*, (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Research, 44(1), pp.344-50.

Asmar, A.T. *et al.*, (2017). Communication across the bacterial cell envelope depends on the size of the periplasm. F. Hughson, ed. PLOS Biology, 15(12), p.e2004303.

Auerswald, E.A., Ludwig, G. & Schaller, H., (1981). Structural analysis of Tn*5*. Cold Spring Harbor symposia on quantitative biology, 45 Pt 1, pp.107-13.

Avissar, Y.J. & Beale, S.I., (1989). Identification of the enzymatic basis for delta-aminolevulinic acid auxotrophy in a *hemA* mutant of *Escherichia coli*. Journal of Bacteriology, 171(6), pp.2919-24.

Baba, T. *et al.*, (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology, (2), 2006.0008.

Badarinarayana, V. *et al.*, (2001). Selection analyses of insertional mutants using subgenic-resolution arrays. Nature Biotechnology, 19(11), pp.1060-1065.

Bakelar, J., Buchanan, S.K. & Noinaj, N., (2016). The structure of the β-barrel assembly machinery complex. Science, 351(6269), pp.180-186.

Bakthavatchalam, Y.D. *et al.*, (2018). Polymyxin susceptibility testing, interpretative breakpoints and resistance mechanisms: An update. Journal of Global Antimicrobial Resistance, 12, pp.124-136.

Bankevich, A. *et al.*, (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology, 19(5), pp.455-77.

Barquist, L. *et al.*, (2013). A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. Nucleic Acids Research, 41(8), pp.4549-64.

Barr, K., Klena, J. & Rick, P.D., (1999). The modality of enterobacterial common antigen polysaccharide chain lengths is regulated by o349 of the *wec* gene cluster of *Escherichia coli* K-12. Journal of Bacteriology, 181(20), pp.6564-8.

Bass, S., Gu, Q. & Christen, A., (1996). Multicopy suppressors of *prc* mutant *Escherichia coli* include two HtrA (DegP) protease homologs (HhoAB), DksA, and a truncated R1pA. Journal of Bacteriology, 178(4), pp.1154-61.

Bendezú, F.O. *et al.*, (2009). RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in *E. coli*. The EMBO Journal, 28(3), pp.193-204.

Bérdy, J., (2012). Thoughts and facts about antibiotics: Where we are now and where we are heading. The Journal of Antibiotics, 65(8), pp.385-395.

Berg, D.E. *et al.*, (1975). Transposition of R factor genes to bacteriophage lambda. Genetics, 72(9), pp.3628-3632.

Berg, D.E., Schmandt, M.A. & Lowe, J.B., (1983). Specificity of transposon Tn*5* insertion. Genetics, 105(4), pp.813-28.

Berglund, N.A. *et al.*, (2015). Interaction of the antimicrobial peptide polymyxin B1 with both membranes of *E. coli*: a molecular dynamics study. D. van der Spoel, ed. PLOS Computational Biology, 11(4), p.e1004180.

Beringer, J.E. *et al.*, (1978). Transfer of the drug-resistance transposon Tn*5* to *Rhizobium*. Nature, 276(5688), pp.633-634.

Bishop, R.E. *et al.*, (2000). Transfer of palmitate from phospholipids to lipid A in outer membranes of Gram-negative bacteria. The EMBO Journal, 19(19), pp.5071-5080.

Bitto, E. & McKay, D.B., (2003). The periplasmic molecular chaperone protein SurA binds a peptide motif that is characteristic of integral outer membrane proteins. The Journal of Biological Chemistry, 278(49), pp.49316-22.

Bladen, H.A. & Mergenhagen, S.E., (1964). Ultrastructure of *Veillonella* and morphological correlation of an outer membrane with particles associated with endotoxic activity. Journal of Bacteriology, 88(5), pp.1482-92.

Blair, J.M. & Piddock, L.J., (2009). Structure, function and inhibition of RND efflux pumps in Gram-negative bacteria: an update. Current Opinion in Microbiology, 12(5), pp.512-519.

Blair, J.M.A. *et al.*, (2015). Molecular mechanisms of antibiotic resistance. Nature Reviews Microbiology, 13(1), pp.42-51.

Bligh, E.G. & Dyer, W.J., (1959). A rapid method of total lipid extraction and purification. Canadian Journal of Biochemistry and Physiology, 37(8), pp.911-917.

Blount, Z.D., (2015). The unexhausted potential of *E. coli*. eLife, 4.

Bolger, A.M., Lohse, M. & Usadel, B., (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), pp.2114-2120.

Brabetz, W. *et al.*, (1997). Deletion of the heptosyltransferase genes *rfaC* and *rfaF* in *Escherichia coli* K-12 results in an Re-type lipopolysaccharide with a high degree of 2-aminoethanol phosphate substitution. European Journal of Biochemistry, 247(2), pp.716-24.

Braun, M. & Silhavy, T.J., (2002). Imp/OstA is required for cell envelope biogenesis in *Escherichia coli.* Molecular Microbiology, 45(5), pp.1289-1302.

Braun, V. & Rehn, K., (1969). Chemical characterization, spatial distribution and function of a lipoprotein (murein-lipoprotein) of the *E. coli* cell wall. The specific effect of trypsin on the membrane structure. European Journal of Biochemistry, 10(3), pp.426-438.

Brochado, A.R. *et al.*, (2018). Species-specific activity of antibacterial drug combinations. Nature, 559(7713), pp.259-263.

Brooks, J.F. *et al.*, (2014). Global discovery of colonization determinants in the squid symbiont *Vibrio fischeri*. Proceedings of the National Academy of Sciences of the United States of America, 111(48), pp.17284-9.

Brown, P.O. & Botstein, D., (1999). Exploring the new world of the genome with DNA microarrays. Nature Genetics, 21(1), pp.33-7.

de Bruijn, F.J., (1987). Transposon Tn*5* mutagenesis to map genes. Methods in Enzymology, 154, pp.175-96.

de Bruijn, F.J. & Lupski, J.R., (1984). The use of transposon Tn*5* mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids--a review. Gene, 27(2), pp.131-49.

Bryan, G., Garza, D. & Hartl, D., (1990). Insertion and excision of the transposable element mariner in *Drosophila*. Genetics, 125(1), pp.103-14.

Bubunenko, M., Baker, T. & Court, D.L., (2007). Essentiality of ribosomal and transcription antitermination proteins analyzed by systematic gene replacement in *Escherichia coli*. Journal of Bacteriology, 189(7), pp.2844-2853.

Byrne, R.T. *et al.*, (2014). *Escherichia coli* genes and pathways involved in surviving extreme exposure to ionizing radiation. Journal of Bacteriology, 196(20), pp.3534-45.

Cabantous, S. & Waldo, G.S., (2006). *In vivo* and *in vitro* protein solubility assays using split GFP. Nature Methods, 3(10), pp.845-854.

Camara, J.E., Skarstad, K. & Crooke, E., (2003). Controlled initiation of chromosomal replication in *Escherichia coli* requires functional Hda protein. Journal of Bacteriology, 185(10), pp.3244-8.

Cano, D.A. *et al.*, (2002). Regulation of capsule synthesis and cell motility in *Salmonella enterica* by the essential gene *igaA*. Genetics, 162(4), pp.1513-23.

Carpenter, C.D. *et al.*, (2014). The Vps/VacJ ABC transporter is required for intercellular spread of *Shigella flexneri.* F. C. Fang, ed. Infection and Immunity, 82(2), pp.660-669.

Casasanta, M.A. *et al.*, (2017). A chemical and biological toolbox for Type Vd secretion: Characterization of the phospholipase A1 autotransporter FplA from *Fusobacterium nucleatum*. The Journal of Biological Chemistry, 292(49), pp.20240-20254.

Ceccarelli, M. & Ruggerone, P., (2008). Physical insights into permeation of and resistance to antibiotics in bacteria. Current Drug Targets, 9(9), pp.779-88.

Chandler, M.G. & Pritchard, R.H., (1975). The effect of gene concentration and relative gene dosage on gene output in *Escherichia coli*. Molecular & General Genetics, 138(2), pp.127-41.

Chang, A.C. & Cohen, S.N., (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic mini-plasmid. Journal of Bacteriology, 134(3), pp.1141-56.

Chao, M.C. *et al.*, (2013). High-resolution definition of the *Vibrio cholerae* essential gene set with hidden Markov model-based analyses of transposon-insertion sequencing data. Nucleic Acids Research, 41(19), pp.9033-48.

Chao, M.C. *et al.*, (2016). The design and analysis of transposon insertion sequencing experiments. Nature Reviews. Microbiology, 14(2), pp.119-28.

Chen, R. & Henning, U., (1996). A periplasmic protein (Skp) of *Escherichia coli* selectively binds a class of outer membrane proteins. Molecular Microbiology, 19(6), pp.1287-94.

Cherepanov, P.P. & Wackernagel, W., (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene, 158(1), pp.9-14.

Chien, A., Edgar, D.B. & Trela, J.M., (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. Journal of Bacteriology, 127(3), pp.1550-1557.

Cho, S.-H. *et al.*, (2014). Detecting envelope stress by monitoring β-Barrel assembly. Cell, 159(7), pp.1652-1664.

Christen, B. *et al.*, (2014). The essential genome of a bacterium. Molecular Systems Biology, 7(1):528.

Chung, B.C. *et al.*, (2013). Crystal Structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. Science, 341(6149), pp.1012-1016.

Claverie-Martin, F. *et al.*, (1991). Analysis of the altered mRNA stability (*ams*) gene from *Escherichia coli*. Nucleotide sequence, transcriptional analysis, and homology of its product to MRP3, a mitochondrial ribosomal protein from *Neurospora crassa*. The Journal of Biological Chemistry, 266(5), pp.2843-51.

Coskun-Ari, F.F. & Hill, T.M., (1997). Sequence-specific interactions in the Tus-Ter complex and the effect of base pair substitutions on arrest of DNA replication in *Escherichia coli*. The Journal of Biological Chemistry, 272(42), pp.26448-56.

Couturier, E. & Rocha, E.P.C., (2006). Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. Molecular Microbiology, 59(5), pp.1506-1518.

Cowan, S.W. *et al.*, (1992). Crystal structures explain functional properties of two *E. coli* porins. Nature, 358(6389), pp.727-733.

Curtis, P.D. & Brun, Y. V, (2014). Identification of essential alphaproteobacterial genes reveals operational variability in conserved developmental and cell cycle systems. Molecular Microbiology, 93(4), pp.713-35.

Dalbey, R.E. & Chen, M., (2004). Sec-translocase mediated membrane protein biogenesis. Biochimica et Biophysica Acta (BBA) Molecular Cell Research, 1694(1–3), pp.37-53.

Dalebroux, Z.D. *et al.*, (2015). Delivery of cardiolipins to the *Salmonella* outer membrane is necessary for survival within host tissues and virulence. Cell Host & Microbe, 17(4), pp.441-51.

Daley, D.O. *et al.*, (2005). Global topology analysis of the *Escherichia coli* inner membrane proteome. Science, 308(5726), pp.1321-1323.

Danese, P.N. *et al.*, (1998). Accumulation of the enterobacterial common antigen lipid II biosynthetic intermediate stimulates *degP* transcription in *Escherichia coli*. Journal of Bacteriology, 180(22), pp.5875-84.

Datsenko, K.A. & Wanner, B.L., (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America, 97(12), pp.6640-6645.

Davies, I.J. & Drabble, W.T., (1996). Stringent and growth-rate-dependent control of the *gua* operon of *Escherichia coli* K-12. Microbiology, 142(9), pp.2429-2437.

Davies, J., (2006). Where have all the antibiotics gone? The Canadian Journal of Infectious Diseases & Medical Microbiology, 17(5), pp.287-290.

Davies, S.C., (2011). Annual Report of the Chief Medical Officer: Volume Two, 2011: Infections and the rise of Antimicrobial Resistance. Department of Health

Deatherage, D.E. & Barrick, J.E., (2014). Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods in Molecular Biology (Clifton, N.J.), 1151, pp.165-188.

DeJesus, M.A. *et al.*, (2017). Statistical analysis of genetic interactions in Tn-Seq data. Nucleic Acids Research, 45(11), p.e93.

DeJesus, M.A. & Ioerger, T.R., (2013). A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. BMC Bioinformatics, 14(1), p.303.

Dembek, M. *et al.*, (2015). High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. mBio, 6(2), p.e02383.

Deris, Z.Z., Akter, J., *et al.*, (2014). A secondary mode of action of polymyxins against Gram-negative bacteria involves the inhibition of NADH-quinone oxidoreductase activity. The Journal of Antibiotics, 67(2), pp.147-151.

Deris, Z.Z., Swarbrick, J.D., *et al.*, (2014). Probing the penetration of antimicrobial polymyxin lipopeptides into Gram-negative bacteria. Bioconjugate Chemistry, 25(4):750-60

Dillon, D.A. *et al.*, (1996). The *Escherichia coli pgpB* gene encodes for a diacylglycerol pyrophosphate phosphatase activity. The Journal of Biological Chemistry, 271(48), pp.30548-30553.

DiRusso, C.C. *et al.*, (1998). Fatty acyl-CoA binding domain of the transcription factor FadR. Characterization by deletion, affinity labeling, and isothermal titration calorimetry. The Journal of Biological Chemistry, 273(50), pp.33652-33659.

Domingues, M.M. *et al.*, (2012). Biophysical characterization of polymyxin B interaction with LPS aggregates and membrane model systems. Biopolymers, 98(4), pp.338-344.

Dorazi, R. & Dewar, S.J., (2000). Membrane topology of the N-terminus of the *Escherichia coli* FtsK division protein. FEBS Letters, 478(1–2), pp.13-18.

Dornenburg, J.E. *et al.*, (2010). Widespread antisense transcription in *Escherichia coli*. mBio, 1(1), pp.e00024-10-e00024-10.

Dowhan, W., (1992). Phosphatidylserine synthase from *Escherichia coli*. Methods in Enzymology, 209, pp.287-298.

Draper, G.C. *et al.*, (1998). Only the N-terminal domain of FtsK functions in cell division. Journal of Bacteriology, 180(17), pp.4621-4627.

Dubarry, N., Possoz, C. & Barre, F.-X., (2010). Multiple regions along the *Escherichia coli* FtsK protein are implicated in cell division. Molecular Microbiology, 78(5), pp.1088-1100.

Duigou, S. *et al.*, (2014). *ssb* gene duplication restores the viability of Δ*holC* and Δ*holD* *Escherichia coli* mutants. D. Hughes, ed. PLoS Genetics, 10(10), p.e1004719.

Duncan, K., van Heijenoort, J. & Walsh, C.T., (1990). Purification and characterization of the D-alanyl-D-alanine-adding enzyme from *Escherichia coli*. Biochemistry, 29(9), pp.2379-2386.

Durand, A. *et al.*, (2016). Mutations affecting potassium import restore the viability of the *Escherichia coli* DNA polymerase III *holD* mutant. I. Matic, ed. PLOS Genetics, 12(6), p.e1006114.

Egan, A.J.F. *et al.*, (2015). Activities and regulation of peptidoglycan synthases. Philosophical transactions of the Royal Society of London. Series B, Biological Sciences, 370(1679).

Eggert, U.S. *et al.*, (2001). Genetic basis for activity differences between vancomycin and glycolipid derivatives of vancomycin. Science (New York, N.Y.), 294(5541), pp.361-364.

Ekiert, D.C. *et al.*, (2017). Architectures of lipid transport systems for the bacterial outer membrane. Cell, 169(2), p.273-285.e17.

Errington, J., (2017). Cell wall-deficient, L-form bacteria in the 21st century: a personal perspective. Biochemical Society Transactions, 45(2), pp.287-295.

Fedoroff, (1995). Biographical Memoirs, Washington, D.C.: National Academies Press.

Fernández, L. *et al.*, (2013). Characterization of the polymyxin B resistome of *Pseudomonas aeruginosa*. Antimicrobial Agents and Chemotherapy, 57(1), pp.110-119.

Fields, S. & Song, O., (1989). A novel genetic system to detect protein–protein interactions. Nature, 340(6230), pp.245-246.

Finn, R.D. *et al.*, (2014). Pfam: the protein families database. Nucleic Acids Research, 42, pp.222-230.

Fiuza, M. *et al.*, (2008). The MurC ligase essential for peptidoglycan biosynthesis is regulated by the serine/threonine protein kinase PknA in *Corynebacterium glutamicum*. Journal of Biological Chemistry, 283(52), pp.36553-36563.

Fleming, A., (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzæ*. British Journal of Experimental Pathology, 10(3), p.226.

Foulquier, E. *et al.*, (2014). PrkC-mediated phosphorylation of overexpressed YvcK protein regulates PBP1 protein localization in *Bacillus subtilis mreB* mutant cells. Journal of Biological Chemistry, 289(34), pp.23662-23669.

Fozo, E.M. *et al.*, (2008). Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. Molecular Microbiology, 70(5), pp.1076-1093.

Freed, N.E., Bumann, D. & Silander, O.K., (2016). Combining *Shigella* Tn-seq data with gold-standard *E. coli* gene deletion data suggests rare transitions between essential and non-essential gene functionality. BMC microbiology, 16(1), p.203.

Frishman, D., (2007). Protein annotation at genomic scale: the current status. Chemical Reviews, 107(8), pp.3448-3466.

Funk, C.R., Zimniak, L. & Dowhan, W., (1992). The *pgpA* and *pgpB* genes of *Escherichia coli* are not essential: evidence for a third phosphatidylglycerophosphate phosphatase. Journal of Bacteriology, 174(1), pp.205-13.

Garwin, J.L., Klages, A.L. & Cronan, J.E., (1980). Beta-ketoacyl-acyl carrier protein synthase II of *Escherichia coli*. Evidence for function in the thermal regulation of fatty acid synthesis. The Journal of Biological Chemistry, 255(8), pp.3263-5.

Gawronski, J.D. *et al.*, (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. Proceedings of the National Academy of Sciences of the United States of America, 106(38), pp.16422-16427.

Georg, J. & Hess, W.R., (2011). cis-antisense RNA, another level of gene regulation in bacteria. Microbiology and Molecular Biology Reviews, 75(2), pp.286-300.

Gerding, M.A. *et al.*, (2007). The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in *E. coli*. Molecular Microbiology, 63(4), pp.1008-1025.

Ghachi, M. El *et al.*, (2005). Identification of multiple genes encoding membrane proteins with undecaprenyl pyrophosphate phosphatase (UppP) activity in *Escherichia coli*. Journal of Biological Chemistry, 280(19), pp.18689-18695.

Girardello, R. *et al.*, (2012). Cation concentration variability of four distinct mueller-hinton agar brands influences polymyxin B susceptibility results. Journal of Clinical Microbiology, 50(7), pp.2414-2418.

Glauert, A.M. & Thornley, M.J., (1969). The topography of the bacterial cell wall. Annual Review of Microbiology, 23(1), pp.159-198.

Glauner, B. & Höltje, J. V, (1990). Growth pattern of the murein sacculus of *Escherichia coli*. The Journal of Biological Chemistry, 265(31), pp.18988-18996.

Goodall, E.C.A. *et al.*, (2018). The essential genome of *Escherichia coli* K-12. S. L. Chen, ed. mBio, 9(1), pp.e02096-17.

Goodman, A.L. *et al.*, (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host & Microbe, 6(3), pp.279-289.

Gould, K., (2016). Antibiotics: from prehistory to the present day. Journal of Antimicrobial Chemotherapy, 71(3), pp.572-575.

Grabowicz, M. & Silhavy, T.J., (2017). Redefining the essential trafficking pathway for outer membrane lipoproteins. Proceedings of the National Academy of Sciences of the United States of America, 114(18), pp.4769-4774.

Grant, A.J. *et al.*, (2016). Genes required for the fitness of *Salmonella enterica* Serovar Typhimurium during infection of immunodeficient gp91 −/− phox mice. A. J. Bäumler, ed. Infection and Immunity, 84(4), pp.989-997.

Green, E.R. & Mecsas, J., (2016). Bacterial secretion systems: an overview. Microbiology Spectrum, 4(1).

Grenov, A.I. & Gerdes, S.Y., (2008). Modeling competitive outgrowth of mutant populations: why do essentiality screens yield divergent results? In Methods in Molecular Biology (Clifton, N.J.). pp. 361-367.

Gunn, J.S. *et al.*, (1998). PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. Molecular Microbiology, 27(6), pp.1171-1182.

Guo, D. & Tropp, B.E., (2000). A second *Escherichia coli* protein with CL synthase activity. Biochimica et Biophysica Acta (BBA) Molecular and Cell Biology of Lipids, 1483(2), pp.263-274.

Gurevich, A. *et al.*, (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics, 29(8), pp.1072-1075.

Han, L. *et al.*, (2016). Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins. Nature Structural & Molecular Biology, 23(3), pp.192-196.

Hanson, A.D. *et al.*, (2009). “Unknown” proteins and “orphan” enzymes: the missing half of the engineering parts list – and how to find it. The Biochemical Journal, 425(1), pp.1-11.

Hantke, K. & Braun, V., (1973). Covalent binding of lipid to protein. Diglyceride and amide-linked fatty acid at the N-terminal end of the murein-lipoprotein of the *Escherichia coli* outer membrane. European Journal of Biochemistry, 34(2), pp.284-296.

Hare, R.S. *et al.*, (2001). Genetic footprinting in bacteria. Journal of Bacteriology, 183(5), pp.1694-706.

Harold, F.M., (1972). Conservation and transformation of energy by bacterial membranes. Bacteriological Reviews, 36(2), pp.172-230.

Harrison, C.J. *et al.*, (1997). Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. Science (New York, N.Y.), 276(5311), pp.431-435.

Harrison, F. *et al.*, (2015). A 1,000-year-old antimicrobial remedy with antistaphylococcal activity. mBio, 6(4), p.e01129.

Harvey, P.C. *et al.*, (2011). *Salmonella enterica* serovar Typhimurium colonizing the lumen of the chicken intestine grows slowly and upregulates a unique set of virulence and metabolism genes. Infection and Immunity, 79(10), pp.4105-4121.

Hassan, K.A. *et al.*, (2016). Fluorescence-based flow sorting in parallel with transposon insertion site sequencing identifies multidrug efflux systems in *Acinetobacter baumannii*. mBio, 7(5), pp.e01200-16.

Hasunuma, K., (2009). Genetics and molecular biology, EOLSS Publishers.

Hayashi, K. *et al.*, (2006). Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Molecular Systems Biology, 2, p.(2006).0007.

Hemm, M.R. *et al.*, (2008). Small membrane proteins found by comparative genomics and ribosome binding site models. Molecular Microbiology, 70(6), pp.1487-501.

Hensel, M. *et al.*, (1995). Simultaneous identification of bacterial virulence genes by negative selection. Science (New York, N.Y.), 269(5222), pp.400-403.

Hiraoka, S., Matsuzaki, H. & Shibuya, I., (1993). Active increase in cardiolipin synthesis in the stationary growth phase and its physiological significance in *Escherichia coli*. FEBS Letters, 336(2), pp.221-224.

Hirota, Y. *et al.*, (1977). On the process of cellular division in *Escherichia coli*: a mutant of *E. coli* lacking a murein-lipoprotein. Proceedings of the National Academy of Sciences of the United States of America, 74(4), pp.1417-1420.

Hobbs, E.C. *et al.*, (2012). Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. Proceedings of the National Academy of Sciences of the United States of America, 109(41), pp.16696-16701.

Holley, R.W. *et al.*, 1965. Structure of ribonucleic acid. Science (New York, N.Y.), 147(3664), pp.1462-1465.

Höltje, J. V, (1998). Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. Microbiology and Molecular Biology Reviews, 62(1), pp.181-203.

Hsu, P.-C. *et al.*, (2017). It is complicated: curvature, diffusion, and lipid sorting within the two membranes of *Escherichia coli*. The Journal of Physical Chemistry Letters, 8(22), pp.5513-5518.

Huang, K.C. *et al.*, (2008). Cell shape and cell-wall organization in Gram-negative bacteria, Proceedings of the National Academy of Sciences of the United States of America, 105(49), pp.19282-19287

Huerta-Cepas, J. *et al.*, (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Molecular Biology and Evolution, 34(8), pp.2115-2122.

Hughes, G.W. *et al.*, (2018). Evidence for phospholipid export from the Gram-negative inner membrane: time to rethink the Mla pathway? bioRxiv, p.388546.

Hulko, M. *et al.*, (2006). The HAMP domain structure implies helix rotation in transmembrane signaling. Cell, 126(5), pp.929-940.

Hulsen, T., de Vlieg, J. & Alkema, W., (2008). BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics, 9(1), p.488.

Hutchison, C.A. *et al.*, (1999). Global transposon mutagenesis and a minimal *Mycoplasma* genome. Science (New York, N.Y.), 286(5447), pp.2165-2169.

Iadanza, M.G. *et al.*, (2016). Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM. Nature Communications, 7, p.12865.

Icho, T., (1988). Membrane-bound phosphatases in *Escherichia coli*: sequence of the *pgpB* gene and dual subcellular localization of the *pgpB* product. Journal of Bacteriology, 170(11), pp.5117-5124.

Icho, T. & Raetz, C.R., (1983). Multiple genes for membrane-bound phosphatases in *Escherichia coli* and their action on phospholipid precursors. Journal of Bacteriology, 153(2), pp.722-730.

Isberg, R.R., Lazaar, A.L. & Syvanen, M., (1982). Regulation of Tn*5* by the right-repeat proteins: control at the level of the transposition reaction? Cell, 30(3), pp.883-892.

Islam, M.S. *et al.*, (2012). Translation of a minigene in the 5′ leader sequence of the enterohaemorrhagic *Escherichia coli* LEE1 transcription unit affects expression of the neighbouring downstream gene. The Biochemical Journal, 441(1), pp.247-253.

Itoh, T. & Tomizawa, J., (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proceedings of the National Academy of Sciences of the United States of America, 77(5), pp.2450-2454.

James, E.S. & Cronan, J.E., (2004). Expression of two *Escherichia coli* acetyl-CoA carboxylase subunits is autoregulated. Journal of Biological Chemistry, 279(4), pp.2520-2527.

Jana, B. *et al.*, (2017). The secondary resistome of multidrug-resistant *Klebsiella pneumoniae*. Scientific Reports, 7(1), p.42483.

Jenks, P. *et al.*, (2001). Identification of nonessential *Helicobacter pylori* genes using random mutagenesis and loop amplification. Research in Microbiology, 152(8), pp.725-734

Jin, T. & Inouye, M., (1994). Transmembrane signalling: mutational analysis of the cytoplasmic linker region of Taz1-1, a Tar-EnvZ chimeric receptor in *Escherichia coli*. Journal of Molecular Biology, 244(5), pp.477-481.

Johnson, D.B.F. *et al.*, (2012). Release factor one is nonessential in *Escherichia coli*. ACS Chemical Biology, 7(8), pp.1337-1344.

Johnson, R.C., Yin, J.C. & Reznikoff, W.S., (1982). Control of Tn*5* transposition in *Escherichia coli* is mediated by protein from the right repeat. Cell, 30(3), pp.873-882.

Jorgenson, M.A. *et al.*, (2016). Dead-end intermediates in the enterobacterial common antigen pathway induce morphological defects in *Escherichia coli* by competing for undecaprenyl phosphate. Molecular Microbiology, 100(1), pp.1-14.

Kamio, Y. & Nikaido, H., (1976). Outer membrane of *Salmonella* Typhimurium: accessibility of phospholipid head groups to phospholipase c and cyanogen bromide activated dextran in the external medium. Biochemistry, 15(12), pp.2561-2570.

Kanazawa, K. *et al.*, (2009). Contribution of each amino acid residue in polymyxin B3 to antimicrobial and lipopolysaccharide binding activity. Chemical & Pharmaceutical Bulletin, 57(3), pp.240-244.

Karimova, G. *et al.*, (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proceedings of the National Academy of Sciences of the United States of America, 95(10), pp.5752-5756.

Katayama, T. *et al.*, (1998). The initiator function of DnaA protein is negatively regulated by the sliding clamp of the *E. coli* chromosomal replicase. Cell, 94(1), pp.61-71.

Kato, J. *et al.*, (1999). The *Escherichia coli* homologue of yeast RER2, a key enzyme of dolichol synthesis, is essential for carrier lipid formation in bacterial cell wall synthesis. Journal of Bacteriology, 181(9), pp.2733-2738.

Kato, J. & Katayama, T., (2001). Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli.* The EMBO Journal, 20(15), pp.4253-4262.

Kawano, M., Aravind, L. & Storz, G., (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. Molecular Microbiology, 64(3), pp.738-754.

Keeney, D., Ruzin. A. & Bradford, P.A., (2007). RamA, a transcriptional regulator, and AcrAB, an RND-type efflux pump, are associated with decreased susceptibility to tigecycline in *Enterobacter cloacae*. Microbial Drug Resistance, 13(1), pp.1-6

Keeney, D. *et al.*, (2008). MarA-mediated overexpression of the AcrAB efflux pump results in decreased susceptibility to tigecycline in *Escherichia coli*. Journal of Antimicrobial Chemotherapy, 61(1), pp.46-53

Kellenberger, E. & Ryter, A., (1958). Cell wall and cytoplasmic membrane of *Escherichia coli.* The Journal of Biophysical and Biochemical Cytology, 4(3), pp.323-326.

Kelley, L.A. *et al.*, (2015). The Phyre2 web portal for protein modelling, prediction and analysis. Nature Protocols, 10(6), pp.845-858.

Keseler, I.M. *et al.*, (2017). The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Research, 45(1), pp.543-550.

Kessels, J.M., Ousen, H. & Van den Bosch, H., (1983). Facilitated utilization of endogenously synthesized lysophosphatidic acid by 1-acylglycerophosphate acyltransferase from *Escherichia coli.* Biochimica et Biophysica Acta, 753(2), pp.227-235.

Khatri, G.S. *et al.*, (1989). The replication terminator protein of *E. coli* is a DNA sequence-specific contra-helicase. Cell, 59(4), pp.667-674.

Kim, S., Grant, R.A. & Sauer, R.T., (2011). Covalent linkage of distinct substrate degrons controls assembly and disassembly of DegP Proteolytic cages. Cell, 145(1), pp.67-78.

Kleckner, N. *et al.*, (1975). Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. Journal of Molecular Biology, 97(4), pp.561-575.

Knowles, T.J. *et al.*, (2009). Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. Nature Reviews Microbiology, 7(3), pp.206-214.

Koboldt, D.C. *et al.*, (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research, 22(3), pp.568-576.

Kolodkin-Gal, I. *et al.*, (2010). D-Amino acids trigger biofilm disassembly. Science, 328(5978), pp.627-629.

Kundig, W., Ghosh, S. & Roseman, S., (1964). Phosphate bound to histidine in a protein as an intermediate in a novel phospho-transferase system. Proceedings of the National Academy of Sciences of the United States of America, 52(4), pp.1067-1074.

Lambden, P.R. & Drabble, W.T., 1973. The *gua* operon of *Escherichia coli* K-12: evidence for polarity from *guaB* to *guaA*. Journal of Bacteriology, 115(3), pp.992-1002.

Langmead, B. & Salzberg, S.L., (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), pp.357-359.

Langridge, G.C. *et al.*, (2009). Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. Genome Research, 19(12), pp.2308-2316.

Larson, T.J. *et al.*, (1980). Membrane phospholipid synthesis in *Escherichia coli*. Identification of the sn-glycerol-3-phosphate acyltransferase polypeptide as the *plsB* gene product. The Journal of Biological Chemistry, 255(19), pp.9421-9426.

Lawyer, F.C. *et al.*, (1993). High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5′ to 3′ exonuclease activity. PCR Methods and Applications, 2(4), pp.275-287.

De Lay, N.R. & Cronan, J.E., (2008). Genetic interaction between the *Escherichia coli* AcpT phosphopantetheinyl transferase and the YejM inner membrane protein. Genetics, 178(3), pp.1327-1337.

Lederberg, J., (1946). Studies in bacterial genetics. Journal of Bacteriology, 52(4), p.503.

Lederberg, J. & Tatum, E.L., (1946). Detection of biochemical mutants of microorganisms. The Journal of Biological Chemistry, 165(1), p.381.

Lee, J. *et al.*, (2016). Characterization of a stalled complex on the β-barrel assembly machine. Proceedings of the National Academy of Sciences of the United States of America, 113(31), pp.8717-8722.

Levine, D.P., (2006). Vancomycin: a history. Clinical Infectious Diseases, 42(1), pp.5-12.

Lewis, K., (2013). Platforms for antibiotic discovery. Nature Reviews Drug Discovery, 12(5), pp.371-387.

Li, G., Hamamoto, K. & Kitakawa, M., (2012). Inner membrane protein YhcB interacts with RodZ involved in cell shape maintenance in *Escherichia coli.* ISRN Molecular Biology, 2012, p.304021.

Li, H. *et al.*, (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), pp.2078-2079.

Li, H. & Durbin, R., (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), pp.1754-1760.

Li, S.J. & Cronan, J.E., (1993). Growth rate regulation of Escherichia coli acetyl coenzyme A carboxylase, which catalyzes the first committed step of lipid biosynthesis. Journal of Bacteriology, 175(2), pp.332-340.

Liebeke, M. *et al.*, (2010). A metabolomic view of *Staphylococcus aureus* and its Ser/Thr kinase and phosphatase deletion mutants: involvement in cell wall biosynthesis. Chemistry & Biology, 17(8), pp.820-830.

Liger, D. *et al.*, (1995). Over-production, purification and properties of the uridine-diphosphate-N-acetylmuramate:L-alanine ligase from *Escherichia coli*. European Journal of Biochemistry, 230(1), pp.80-87.

Lima, S. *et al.*, (2013). Dual molecular signals mediate the bacterial response to outer-membrane stress. Science, 340(6134), pp.837-841.

Liu, A. *et al.*, (2010). Antibiotic sensitivity profiles determined with an *Escherichia coli* gene knockout collection: generating an antibiotic bar code. Antimicrobial Agents and Chemotherapy, 54(4), pp.1393-1403.

Liu, D. & Reeves, P.R., (1994). *Escherichia coli* K-12 regains its O-antigen. Microbiology, 140(1), pp.49-57.

Liu, G., Draper, G.C. & Donachie, W.D., (1998). FtsK is a bifunctional protein involved in cell division and chromosome localization in *Escherichia coli*. Molecular Microbiology, 29(3), pp.893-903.

Liu, Y. & Breukink, E., (2016). The membrane steps of bacterial cell wall synthesis as antibiotic targets. Antibiotics, 5(3), p.28.

Lodge, J. *et al.*, (1992). Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. FEMS Microbiology Letters, 95(2–3), pp.271-276.

de Lorenzo, V. *et al.*, (1990).. Mini-Tn*5* transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. Journal of Bacteriology, 172(11), pp.6568-6572.

Louie, K. & Dowhan, W., (1980). Investigations on the association of phosphatidylserine synthase with the ribosomal component from *Escherichia coli*. The Journal of Biological Chemistry, 255(3), pp.1124-1127.

Lu, Y.-H. *et al.*, (2011). Three phosphatidylglycerol-phosphate phosphatases in the inner membrane of *Escherichia coli*. Journal of Biological Chemistry, 286(7), pp.5506-5518.

Luirink, J. *et al.*, (2012). Biogenesis of inner membrane proteins in *Escherichia coli*. Biochimica et Biophysica Acta (BBA) Bioenergetics, 1817(6), pp.965-976.

Maddalo, G. *et al.*, (2011). Systematic analysis of native membrane protein complexes in *Escherichia coli*. Journal of Proteome Research, 10(4), pp.1848-1859.

Mahoney, T.F., Ricci, D.P. & Silhavy, T.J., (2016). Classifying β-barrel Assembly substrates by manipulating essential Bam complex members. P. J. Christie, ed. Journal of Bacteriology, 198(14), pp.1984-1992.

Malinverni, J.C. & Silhavy, T.J., (2009). An ABC transport system that maintains lipid asymmetry in the Gram-negative outer membrane. Proceedings of the National Academy of Sciences, 106(19), pp.8009-8014.

Mann, B. *et al.*, (2012). Control of virulence by small RNAs in *Streptococcus pneumoniae.* P. Cossart, ed. PLoS Pathogens, 8(7), p.e1002788.

Manna, D. *et al.*, (2007). Microarray analysis of Mu transposition in *Salmonella enterica*, serovar Typhimurium: transposon exclusion by high-density DNA binding proteins. Molecular Microbiology, 66(2), pp.315-328.

Manoil, C. & Beckwith, J., (1985). Tn*phoA*: a transposon probe for protein export signals. Proceedings of the National Academy of Sciences of the United States of America, 82(23), pp.8129-8133.

Mares, J. *et al.*, (2009). Interactions of lipopolysaccharide and polymyxin studied by NMR spectroscopy. The Journal of Biological Chemistry, 284(17), pp.11498-11506.

Marquardt, J.L. *et al.*, (1992). Cloning and sequencing of *Escherichia coli murZ* and purification of its product, a UDP-N-acetylglucosamine enolpyruvyl transferase. Journal of Bacteriology, 174(17), pp.5748-5752.

Martorana, A.M. *et al.*, (2011). Complex transcriptional organization regulates an *Escherichia coli* locus implicated in lipopolysaccharide biogenesis. Research in Microbiology, 162(5), pp.470-482.

Masi, M & Pagès, J.M., (2013). Structure, Function and Regulation of Outer Membrane Proteins Involved in Drug Transport in Enterobacteriacea: the OmpF/C - TolC Case. Open Microbiology Journal, 7, pp.22-33

Matsuyama, S., Tajima, T. & Tokuda, H., (1995). A novel periplasmic carrier protein involved in the sorting and transport of *Escherichia coli* lipoproteins destined for the outer membrane. The EMBO Journal, 14(14), pp.3365-3372.

May, K.L. & Silhavy, T.J., (2018). The *Escherichia coli* phospholipase PldA regulates outer membrane homeostasis via lipid signalling. S. Gottesman, ed. mBio, 9(2).

Mazodier, P. *et al.*, (1985). Completion of the nucleotide sequence of the central region of Tn*5* confirms the presence of three resistance genes. Nucleic Acids Research, 13(1), pp.195-205.

McAteer, S. *et al.*, (2001). The *lytB* gene of *Escherichia coli* is essential and specifies a product needed for isoprenoid biosynthesis. Journal of Bacteriology, 183(24), pp.7403-7407.

McClintock, B., (1950). The origin and behaviour of mutable loci in maize. Proceedings of the National Academy of Sciences of the United States of America, 36(6), pp.344-355.

McGrath, B.C. & Osborn, M.J., (1991). Localization of the terminal steps of O-antigen synthesis in *Salmonella* Typhimurium. Journal of Bacteriology, 173(2), pp.649-654.

Mehdizadeh Aghdam, E., Hejazi, M.S. & Barzegar, A., (2016). Riboswitches: From living biosensors to novel targets of antibiotics. Gene, 592(2), pp.244-259.

Mei, J.M. *et al.*, (1997). Identification of *Staphylococcus aureus* virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. Molecular Microbiology, 26(2), pp.399-407.

Mendoza-Vargas, A. *et al.*, (2009). Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli.* C. Creighton, ed. PLoS ONE, 4(10), p.e7526.

Mengin-Lecreulx, D. *et al.*, (1991). The *murG* gene of *Escherichia coli* codes for the UDP-N-acetylglucosamine: N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis. Journal of Bacteriology, 173(15), pp.4625-4636.

Meredith, T.C. *et al.*, (2012). Harnessing the power of transposon mutagenesis for antibacterial target identification and evaluation. Mobile Genetic Elements, 2(4), pp.171-178.

Mi, H. *et al.*, (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Research, 45(1), pp.183-189.

Mileykovskaya, E. & Dowhan, W., (2000). Visualization of phospholipid domains in escherichia coli by using the cardiolipin-specific fluorescent dye 10-N-Nonyl Acridine Orange. Journal of Bacteriology, 182(4), pp.1172-1175.

Min Jou, W. *et al.*, (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. Nature, 237(5350), pp.82-88.

Moffatt, J.H. *et al.*, (2010). Colistin resistance in *Acinetobacter baumannii* is mediated by complete loss of lipopolysaccharide production. Antimicrobial Agents and Chemotherapy, 54(12), pp.4971-4977.

Mogi, T. *et al.*, (2006). Role of a putative third subunit YhcB on the assembly and function of cytochrome bd-type ubiquinol oxidase from *Escherichia coli*. Biochimica et Biophysica Acta (BBA) Bioenergetics, 1757(7), pp.860-864.

Morgenstein, R.M. *et al.*, (2015). RodZ links MreB to cell wall synthesis to mediate MreB rotation and robust morphogenesis. Proceedings of the National Academy of Sciences of the United States of America, 112(40), pp.12510-12515.

Mühlradt, P.F. & Golecki, J.R., (1975). Asymmetrical distribution and artifactual reorientation of lipopolysaccharide in the outer membrane bilayer of *Salmonella* Typhimurium. European Journal of Biochemistry, 51(2), pp.343-352.

Murakami, A., Nakatogawa, H. & Ito, K., (2004). Translation arrest of SecM is essential for the basal and regulated expression of SecA. Proceedings of the National Academy of Sciences of the United States of America, 101(33), pp.12330-12335.

My, L. *et al.*, (2015). Reassessment of the genetic regulation of fatty acid synthesis in escherichia coli: global positive control by the dual functional regulator FadR. P. de Boer, ed. Journal of Bacteriology, 197(11), pp.1862-1872.

Nguyen, B.D. & Valdivia, R.H., (2012). Virulence determinants in the obligate intracellular pathogen *Chlamydia trachomatis* revealed by forward genetic approaches. Proceedings of the National Academy of Sciences of the United States of America, 109(4), pp.1263-1268.

Nichols, R.J. *et al.*, (2011). Phenotypic landscape of a bacterial cell. Cell, 144(1), pp.143-156.

Nikaido, H., (2005). Restoring permeability barrier function to outer membrane. Chemistry & Biology, 12(5), pp.507-509.

Nishijima, S. *et al.*, (1988). Disruption of the *Escherichia coli cls* gene responsible for cardiolipin synthesis. Journal of Bacteriology, 170(2), pp.775-780.

Noinaj, N. *et al.*, (2014). Lateral opening and exit pore formation are required for BamA function. Structure, 22(7), pp.1055-1062.

Okusu, H., Ma, D. & Nikaido, H., (1996). AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants. Journal of Bacteriology, 178(1), pp.306-8.

Olaitan, A.O., Morand, S. & Rolain, J.-M., (2014). Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. Frontiers in Microbiology, 5, p.643.

Opdyke, J.A., Kang, J.-G. & Storz, G., (2004). GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. Journal of Bacteriology, 186(20), pp.6698-6705.

van Opijnen, T., Bodi, K.L. & Camilli, A., (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nature Methods, 6(10), pp.767-772.

van Opijnen, T. & Camilli, A., (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nature Reviews. Microbiology, 11(7), pp.435-442.

Osawa, K. *et al.*, (2013). Modulation of O-antigen chain length by the *wzz* gene in *Escherichia coli* O157 influences its sensitivities to serum complement. Microbiology and Immunology, 57(9), pp.616-623

Ow, M.C. *et al.*, (2002). RNase E levels in *Escherichia coli* are controlled by a complex regulatory system that involves transcription of the *rne* gene from three promoters. Molecular Microbiology, 43(1), pp.159-171.

Pan, X. *et al.*, (2006). A DNA integrity network in the yeast *Saccharomyces cerevisiae*. Cell, 124(5), pp.1069-1081.

Papanastasiou, M. *et al.*, (2013). The *Escherichia coli* peripheral inner membrane proteome. Molecular & Cellular Proteomics, 12(3), p.599.

Parsons, A.B. *et al.*, (2004). Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. Nature Biotechnology, 22(1), pp.62-69.

Paulsen, I.T., Cain, A.K. & Hassan, K.A., (2017). Physical enrichment of transposon mutants from saturation mutant libraries using the TraDISort approach. Mobile Genetic Elements, 7(3), pp.1-7.

Paulus, H. & Gray, E., (1964). The biosynthesis of polymyxin B by growing cultures of *Bacillus polymyxa*. The Journal of Biological Chemistry, 239, pp.865-871.

Pearson, W.R. *et al.*, (1997). Comparison of DNA sequences with protein sequences. Genomics, 46(1), pp.24-36.

Peng, D. *et al.*, (2005). *Moraxella catarrhalis* bacterium without endotoxin, a potential vaccine candidate. Infection and Immunity, 73(11), pp.7569-7577.

Peterson, A.A., Hancock, R.E. & McGroarty, E.J., (1985). Binding of polycationic antibiotics and polyamines to lipopolysaccharides of *Pseudomonas aeruginosa*. Journal of Bacteriology, 164(3), pp.1256-1261.

Phan, M.D. *et al.*, (2015). Molecular characterization of a multidrug resistance IncF plasmid from the globally disseminated *Escherichia coli* ST131 clone. PloS one, 10(4), p.e0122369.

Phan, M.D. *et al.*, (2013). The serum resistome of a globally disseminated multidrug resistant uropathogenic *Escherichia coli* clone. PLoS Genetics, 9(10), p.e1003834.

Piggot, T.J., Holdbrook, D.A. & Khalid, S., (2011). Electroporation of the *E. coli* and *S. aureus* membranes: molecular dynamics simulations of complex bacterial membranes. The Journal of Physical Chemistry B, 115(45), pp.13381-13388.

Pluschke, G., Hirota, Y. & Overath, P., (1978). Function of phospholipids in *Escherichia coli*. Characterization of a mutant deficient in cardiolipin synthesis. The Journal of Biological Chemistry, 253(14), pp.5048-55.

Poldermans, B., Roza, L. & Van Knippenberg, P.H., (1979). Studies on the function of two adjacent N6,N6-dimethyladenosines near the 3' end of 16 S ribosomal RNA of *Escherichia coli*. III. Purification and properties of the methylating enzyme and methylase-30 S interactions. The Journal of Biological Chemistry, 254(18), pp.9094-9100.

Polissi, A. & Sperandeo, P., (2014). The lipopolysaccharide export pathway in *Escherichia coli*: structure, organization and regulated assembly of the Lpt machinery. Marine Drugs, 12(2), pp.1023-1042.

Poltorak, A. *et al.*, (1998). Defective LPS signalling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science (New York, N.Y.), 282(5396), pp.2085-2088.

Pomerantz, R.T. & O'Donnell, M., (2010). What happens when replication and transcription complexes collide? Cell cycle (Georgetown, Tex.), 9(13), pp.2537-2543.

Quinlan, A.R. & Hall, I.M., (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), pp.841-842.

Raetz, C.R.H. & Whitfield, C., (2002). Lipopolysaccharide endotoxins. Annual Review of Biochemistry, 71, pp.635-700.

Raghavan, R., Sloan, D.B. & Ochman, H., (2012). Antisense transcription is pervasive but rarely conserved in enteric bacteria. mBio, 3(4), e00156-12.

Ramos-Morales, F. *et al.*, (2003). Role for *Salmonella enterica* enterobacterial common antigen in bile resistance and virulence. Journal of Bacteriology, 185(17), pp.5328-5332.

Ravindran, S., (2012). Barbara McClintock and the discovery of jumping genes. Proceedings of the National Academy of Sciences of the United States of America, 109(50), pp.20198-20199.

Reich, K.A., Chovan, L. & Hessler, P., (1999). Genome scanning in *Haemophilus influenzae* for identification of essential genes. Journal of Bacteriology, 181(16), pp.4961-4968.

Renner, L.D. & Weibel, D.B., (2011). Cardiolipin microdomains localize to negatively curved regions of *Escherichia coli* membranes. Proceedings of the National Academy of Sciences of the United States of America, 108(15), pp.6264-6269.

Reyes-Lamothe, R., Sherratt, D.J. & Leake, M.C., (2010). Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. Science (New York, N.Y.), 328(5977), pp.498-501.

Reznikoff, W.S., (1993). The Tn*5* transposon, Riber, L. *et al.*, (2006). Hda-mediated inactivation of the DnaA protein and *dnaA* gene autoregulation act in concert to ensure homeostatic maintenance of the *Escherichia coli* chromosome. Genes & Development, 20(15), pp.2121-2134.

Rice, L.B., (2010). Progress and challenges in implementing the research on ESKAPE pathogens. Infection Control & Hospital Epidemiology, 31(1), pp.7-10.

Riley, M. *et al.*, (2006). *Escherichia coli* K-12: a cooperatively developed annotation snapshot--(2005). Nucleic Acids Research, 34(1), pp.1-9.

Rizzitello, A.E., Harper, J.R. & Silhavy, T.J., (2001). Genetic evidence for parallel pathways of chaperone activity in the periplasm of *Escherichia coli.* Journal of Bacteriology, 183(23), pp.6794-800.

Robinson, A., (2016). The development and application of a transposon insertion sequencing methodology in *Escherichia coli* BW25113. Ph.D. thesis, College of Medical and Dental Sciences, University of Birmingham

Rocha, E.P.C. & Danchin, A., (2003). Gene essentiality determines chromosome organisation in bacteria. Nucleic Acids Research, 31(22), pp.6570-6577.

Rothstein, S.J. *et al.*, (1980). The inverted repeats of Tn*5* are functionally different. Cell, 19(3), pp.795-805.

Rothstein, S.J. & Reznikoff, W.S., (1981). The functional differences in the inverted repeats of Tn*5* are caused by a single base pair nonhomology. Cell, 23(1), pp.191-199.

Rowlett, V.W. *et al.*, (2017). Impact of membrane phospholipid alterations in *Escherichia coli* on cellular function and bacterial stress adaptation. Journal of Bacteriology, 199(13), pp.e00849-16.

Ruiz, N. *et al.*, (2005). Chemical conditionality: a genetic strategy to probe organelle assembly. Cell, 121(2), pp.307-317.

Rutherford, K. *et al.*, (2000). Artemis: sequence visualization and annotation. Bioinformatics (Oxford, England), 16(10), pp.944-945.

Sanger, F. *et al.*, (1977). Nucleotide sequence of bacteriophage phi X174 DNA. Nature, 265(5596), pp.687-695.

Sanger, F., Brownlee, G.G. & Barrell, B.G., (1965). A two-dimensional fractionation procedure for radioactive nucleotides. Journal of Molecular Biology, 13(2), pp.373-398.

Sankaran, K. & Wu, H.C., (1994). Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. The Journal of Biological Chemistry, 269(31), pp.19701-19706.

Santa Maria, J.P. *et al.*, (2014). Compound-gene interaction mapping reveals distinct roles for *Staphylococcus aureus* teichoic acids. Proceedings of the National Academy of Sciences of the United States of America, 111(34), pp.12510-12515.

Santajit, S. & Indrawattana, N., (2016). Mechanisms of antimicrobial resistance in ESKAPE pathogens. BioMed Research International, 2016, p.2475067.

Santiago, M. *et al.*, (2015). A new platform for ultra-high density *Staphylococcus aureus* transposon libraries. BMC Genomics, 16(1), p.252.

Sarmiento, F., Mrázek, J. & Whitman, W.B., (2013). Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. Proceedings of the National Academy of Sciences of the United States of America, 110(12), pp.4726-4731.

Sassetti, C.M., Boyd, D.H. & Rubin, E.J., (2001). Comprehensive identification of conditionally essential genes in mycobacteria. Proceedings of the National Academy of Sciences of the United States of America, 98(22), pp.12712-12717.

Sassetti, C.M., Boyd, D.H. & Rubin, E.J., (2003). Genes required for mycobacterial growth defined by high density mutagenesis. Molecular Microbiology, 48(1), pp.77-84.

Sauert, M. *et al.*, (2016). The MazF-regulon: a toolbox for the post-transcriptional stress response in *Escherichia coli*. Nucleic Acids Research, 44(14), pp.6660-6675.

Schneck, E. *et al.*, (2010). Quantitative determination of ion distributions in bacterial lipopolysaccharide membranes by grazing-incidence X-ray fluorescence. Proceedings of the National Academy of Sciences, 107(20), pp.9147-9151.

Schulz, G.E., (1993). Bacterial porins: structure and function. Current Opinion in Cell Biology, 5(4), pp.701-707.

Schulz, G.E., (2002). The structure of bacterial outer membrane proteins. Biochimica et biophysica acta, 1565(2), pp.308-317.

Schwechheimer, C., Rodriguez, D.L. & Kuehn, M.J., (2015). NlpI-mediated modulation of outer membrane vesicle production through peptidoglycan dynamics in *Escherichia coli*. Microbiology Open, 4(3), pp.375-389.

Shah, I.M. *et al.*, (2008). A Eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. Cell, 135(3), pp.486-496.

Shaikh, S. *et al.*, (2015). Antibiotic resistance and extended spectrum beta-lactamases: Types, epidemiology and treatment. Saudi Journal of Biological Sciences, 22(1), pp.90-101.

Silhavy, T.J., Kahne, D. & Walker, S., (2010). The bacterial cell envelope. Cold Spring Harbor Perspectives in Biology, 2(5), p.a000414.

Silver, L.L., (2011). Challenges of antibacterial discovery. Clinical Microbiology Reviews, 24(1), pp.71-109.

Simpson, B.W. *et al.*, (2015). Lipopolysaccharide transport to the cell surface: biosynthesis and extraction from the inner membrane. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 370(1679):26370941.

Singh, S.K. *et al.*, (2015). Regulated proteolysis of a cross-link-specific peptidoglycan hydrolase contributes to bacterial morphogenesis. Proceedings of the National Academy of Sciences of the United States of America, 112(35), pp.10956-10961.

Singh, S.K. *et al.*, (2012). Three redundant murein endopeptidases catalyse an essential cleavage step in peptidoglycan synthesis of *Escherichia coli* K12. Molecular Microbiology, 86(5), pp.1036-1051.

Singh, S.S. *et al.*, (2014). Widespread suppression of intragenic transcription initiation by H-NS. Genes & development, 28(3), pp.214-219.

Sklar, J.G. *et al.*, (2007). Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. Genes & Development, 21(19), pp.2473-2484.

Smith, V. *et al.*, (1996). Functional analysis of the genes of yeast chromosome V by genetic footprinting. Science (New York, N.Y.), 274(5295), pp.2069-2074.

Smith, V., Botstein, D. & Brown, P.O., (1995). Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. Proceedings of the National Academy of Sciences of the United States of America, 92(14), pp.6479-6483.

Sneppen, K. *et al.*, (2005). A Mathematical Model for Transcriptional Interference by RNA Polymerase Traffic in *Escherichia coli*. Journal of Molecular Biology, 346(2), pp.399-409.

Solaimanpour, S., Sarmiento, F. & Mrázek, J., (2015). Tn-seq explorer: a tool for analysis of high-throughput sequencing data of transposon mutant libraries. PloS One, 10(5), p.e0126070.

Sperandeo, P. *et al.*, (2006). Non-essential KDO biosynthesis and new essential cell envelope biogenesis genes in the *Escherichia coli yrbG–yhbG* locus. Research in Microbiology, 157(6), pp.547-558.

Spiess, C., Beil, A. & Ehrmann, M., (1999). A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. Cell, 97(3), pp.339-347.

Srimal, S. *et al.*, (1996). Titration calorimetric studies to elucidate the specificity of the interactions of polymyxin B with lipopolysaccharides and lipid A. The Biochemical Journal, 315 (Pt 2)(Pt 2), pp.679-686.

Srivatsan, A. *et al.*, (2010). Co-Orientation of replication and transcription preserves genome integrity N. A. Moran, ed. PLoS Genetics, 6(1), p.e1000810.

Steeghs, L. *et al.*, (1998). Meningitis bacterium is viable without endotoxin. Nature, 392(6675), pp.449-449.

Stenberg, F. *et al.*, (2005). Protein complexes of the *Escherichia coli* cell envelope. Journal of Biological Chemistry, 280(41), pp.34409-34419.

Stevenson, G. *et al.*, (1994). Structure of the O-antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. Journal of Bacteriology, 176(13), pp.4144-4156.

Subashchandrabose, S. *et al.*, (2016). *Acinetobacter baumannii* genes required for bacterial survival during bloodstream infection. mSphere, 1(1), e00013-15.

Sutterlin, H.A. *et al.*, (2016). Disruption of lipid homeostasis in the Gram-negative cell envelope activates a novel cell death pathway. Proceedings of the National Academy of Sciences of the United States of America, 113(11), pp.1565-1574.

Suzuki, H. *et al.*, (1978). Murein-lipoprotein of *Escherichia coli*: a protein involved in the stabilization of bacterial cell envelope. Molecular & General Genetics, 167(1), pp.1-9.

Tan, B.K. *et al.*, (2012). Discovery of a cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as substrates. Proceedings of the National Academy of Sciences, 109(41), pp.16504-16509.

Tatusov, R.L. *et al.*, (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research, 28(1), pp.33-36.

Taylor, A.L. & Dunham Trottfer, C., (1967). Revised Linkage Map of *Escherichia coli,* Bacteriology Review 31:332-353

Taylor, A.L. & Thoman, M.S., (1964). The genetic map of *Escherichia coli* K-12. Genetics, 50, pp.659-677.

Teuber, M. & Bader, J., (1976). Action of polymyxin B on bacterial membranes. Binding capacities for polymyxin B of inner and outer membranes isolated from *Salmonella* Typhimurium G30. Archives of Microbiology, 109(1), pp.51-8.

The Gene Ontology Consortium, (2017). Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Research, 45(1), pp.331-338.

Thomason, L.C., Costantino, N. & Court, D.L., (2007). *E. coli* Genome Manipulation by P1 Transduction. In Current Protocols in Molecular Biology. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp.1.17.1-8.

Thomason, M.K. *et al.*, (2015). Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. Journal of Bacteriology, 197(1), pp.18-28.

Thong, S. *et al.*, (2016). Defining key roles for auxiliary proteins in an ABC transporter that maintains bacterial outer membrane lipid asymmetry. eLife, 5, e19042.

Thorpe, H.A. *et al.*, (2016). The large majority of intergenic sites in bacteria are selectively constrained, even when known regulatory elements are excluded. bioRxiv, p.069708.

Tjaden, B. *et al.*, (2002). Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. Nucleic Acids Research, 30(17), pp.3732-3738.

Tong, X. *et al.*, (2004). Genome-scale identification of conditionally essential genes in *E. coli* by DNA microarrays. Biochemical and Biophysical Research Communications, 322(1), pp.347-354.

Trimble, J.M. *et al.*, (2016). Polymyxin: alternative mechanisms of action and resistance. CSH Perspectives, 6(10) a025288

Troy, E.B. *et al.*, (2016). Global Tn-seq analysis of carbohydrate utilization and vertebrate infectivity of *Borrelia burgdorferi*. Molecular Microbiology, 101(6), pp.1003-1023.

Typas, A. *et al.*, (2008). High-throughput, quantitative analyses of genetic interactions in *E. coli*. Nature Methods, 5(9), pp.781-787.

Uehara, T. & Park, J.T., (2008). Growth of *Escherichia coli*: significance of peptidoglycan degradation during elongation and septation. Journal of Bacteriology, 190(11), pp.3914-3922.

Vaara, M. & Vaara, T., (1983). Sensitization of Gram-negative bacteria to antibiotics and complement by a nontoxic oligopeptide. Nature, 303(5917), pp.526-528.

Velkov, T. *et al.*, (2013). Pharmacology of polymyxins: new insights into an "old" class of antibiotics. Future Microbiology, 8(6), pp.711-724.

Vertommen, D. *et al.*, (2009). Characterization of the role of the *Escherichia coli* periplasmic chaperone SurA using differential proteomics. Proteomics, 9(9), pp.2432-2443.

Viguera, E. *et al.*, (2003). Lethality of bypass polymerases in *Escherichia coli* cells with a defective clamp loader complex of DNA polymerase III. Molecular Microbiology, 50(1), pp.193-204.

Wainwright, M., (1989). Moulds in ancient and more recent medicine. Mycologist, 3(1), pp.21-23.

Waller, P.R. & Sauer, R.T., (1996). Characterization of *degQ* and *degS*, *Escherichia coli* genes encoding homologs of the DegP protease. Journal of Bacteriology, 178(4), pp.1146-1153.

Wang, L. & Lutkenhaus, J., (1998). FtsK is an essential cell division protein that is localized to the septum and induced as part of the SOS response. Molecular Microbiology, 29(3), pp.731-740.

Weimar, J.D. *et al.*, (2002). Functional role of fatty acyl-Coenzyme A synthetase in the transmembrane movement and activation of exogenous long-chain fatty acids. Journal of Biological Chemistry, 277(33), pp.29369-29376.

WHO, (2018). Antibiotic resistance. WHO. Available at: http://www.who.int/en/news-room/fact-sheets/detail/antibiotic-resistance [Accessed August 2, 2018].

Williams, K.J., (2009). The introduction of "chemotherapy" using arsphenamine – the first magic bullet. Journal of the Royal Society of Medicine, 102(8), pp.343-348.

Wilson, J.W. *et al.*, (2007). Space flight alters bacterial gene expression and virulence and reveals a role for global regulator Hfq. Proceedings of the National Academy of Sciences of the United States of America, 104(41), pp.16299-16304.

Wolf, J., Gerber, A.P. & Keller, W., (2002). *tadA*, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. The EMBO Journal, 21(14), pp.3841-3851.

Wollman, E.L. & Jacob, F., (1955). [Mechanism of the transfer of genetic material during recombination in Escherichia coli K12]. Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences, 240(25), pp.2449-2451.

Wong, S.M. & Mekalanos, J.J., (2000). Genetic footprinting with mariner-based transposition in *Pseudomonas aeruginosa*. Proceedings of the National Academy of Sciences of the United States of America, 97(18), pp.10191-10196.

Wu, T. *et al.*, (2006). Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of *Escherichia coli*. Proceedings of the National Academy of Sciences, 103(31), pp.11754-11759.

Yamaguchi, K., Yu, F. & Inouye, M., (1988). A single amino acid determinant of the membrane localization of lipoproteins in *E. coli*. Cell, 53(3), pp.423-432.

Yamamoto, N. *et al.*, (2009). Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. Molecular Systems Biology, (5)335.

Yamazaki, Y., Niki, H. & Kato, J., (2008). Profiling of *Escherichia coli* Chromosome database. Methods in Molecular Biology (Clifton, N.J.), 416, pp.385-389.

Yu, X.C. *et al.*, (1998). Localization of cell division protein FtsK to the Escherichia coli septum and identification of a potential N-terminal targeting domain. Journal of Bacteriology, 180(5), pp.1296-1304.

Zavascki, A.P. *et al.*, (2007). Polymyxin B for the treatment of multidrug-resistant pathogens: a critical review. Journal of Antimicrobial Chemotherapy, 60(6), pp.1206-1215.

Zhang, Y.J. *et al.*, (2012). Global assessment of genomic regions required for growth in Mycobacterium tuberculosis. PLoS Pathogens, 8(9), p.e1002946.

Zhou, J. & Rudd, K.E., (2013). EcoGene 3.0. Nucleic Acids Research, 41, pp.613-624.

Zhou, Z. *et al.*, (1998). Function of Escherichia coli MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis. The Journal of Biological Chemistry, 273(20), pp.12466-12475.

Zhu, L. *et al.*, (2010). Triclosan resistance of *Pseudomonas aeruginosa* PAO1 is due to FabV, a triclosan-resistant enoyl-acyl carrier protein reductase. Antimicrobial Agents and Chemotherapy, 54(2), pp.689-698.

Zomer, A. *et al.*, (2012). ESSENTIALS: software for rapid analysis of high throughput transposon insertion sequencing data. PloS One, 7(8), p.e43012.

Zückert, W.R., (2014). Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. Biochimica et Biophysica Acta, 1843(8), pp.1509-1516.

Zuker, M., (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research, 31(13), pp.3406-3415.